Regarding the Claim Objections

Claims 22 to 24 stand objected to due to the dependency of the claims being incorrect. As set forth herein, claims 22 to 24 have been amended to provide the correct claim dependency. As such, withdrawal of the objection is respectfully requested.

I. DOUBLE PATENTING REJECTION

The rejection of claims 25 to 28 is respectfully traversed. The Examiner indicates that these claims are substantial duplicates of claims 6 to 9.

As set forth herein, claims 25 to 28 have been amended to depend from claim 21 thereby providing the correct claim dependency. Thus, amended claims 25 to 28 are not substantial duplicates of claims 6 to 9 and, as such, withdrawal of the rejection is respectfully requested.

II. REJECTION UNDER 35 U.S.C. §112

The rejection of claims 1 to 9 and 21 to 28 under 35 U.S.C. §112, first paragraph, as allegedly lacking an adequate written description, is respectfully traversed. The Examiner states that "given the enormous scope and divergent nature of the subject matter encompassed by the Factor IX of these claims, these examples fail to adequately represent the full genus." [page 6, first paragraph of the Office Action] *Vas-Cath Inc. v. Mahurakar* (935 F.2d 1555 (Fed. Cir. 1991)), *Fiers v. Revel* (984 F.2d 1164 (Fed. Cir. 1993)) and *Reagents of the Univ. Calif. v. Eli Lilly* (119 F.3d 1559 (Fed. Cir. 1997)) are cited in support of the rejection.

Claims 1 to 9 and 21 to 28 are adequately described. To provide an adequate written description for gene sequences "requires more than a mere statement that it is part of the invention ....what is required is a description of the DNA itself." *Fiers* at 1170-71. In *Reagents of the Univ. Calif. v. Eli Lilly,* the Federal Circuit explained that "description of a genus of cDNAs may be achieved by means of a recitation of a representative number of cDNAs, defined by nucleotide sequence, falling within the scope of the genus or of a recitation of structural features common to the members of the genus, which constitute a substantial portion of the genus." *Id.* at 1568. Although the *Lilly* court did not specify how many species constitutes a representative number, "Applicants are not required to disclose every species encompassed by

their claims, even in an unpredictable art." *In re Angstadt*, 537 F.2d, 498, 502-503 (CCPA 1976).

Here, the specification discloses a representative number of Factor IX species to adequately describe a Factor IX genus. Furthermore, knowledge in the art regarding Factor IX structure and function, as well as critical and non-critical amino acid sequences for Factor IX function, was extensive at the time of the invention.

First, the Examiner acknowledges that the specification discloses five mammalian Factor IX species, human, mouse, canine, rabbit and bovine. The Examiner also acknowledges that the specification discloses a Factor IX variant having function. [see, page 6, first paragraph of the Office Action] Thus, the specification discloses a total of six different mammalian Factor IX species.

Second, knowledge in the art regarding the structure and function of Factor IX was extensive at the time of the invention. To corroborate Applicant's position, submitted herewith as Exhibits A to D are references by Sarkar *et al.* (Genomics 6:133 (1990)); Kurachi *et al.* (Blood Coagulation and Fibrinolysis 4:953 (1993)); Bottema *et al.* (Am. J. Hum. Genet. 49:820 (1991)), and Giannelli *et al.* (Nucleic Acids Res. 20:2027 (1992)).

In Exhibit A, the authors describe sequencing the activation peptide and catalytic domain of Factor IX in six species, sheep, pig, rabbit, guinea pig, rat and mouse. Thus, mammalian Factor IX sequences, in addition to those referred to in the specification, were known in the art at the time of the invention.

In Exhibit B, the detailed structure and function of Factor IX regions is described (pages 954-958). In particular, for example, Factor IX domains are illustrated in Fig. 2 (page 955, catalytic, GLA, EGF domains, etc.), conserved regions described (propeptide, page 955, second column), domains important for function described (Gal domain Ca^{++} binding, page 956, paragraph bridging the first and second columns; EGF domain, page 957, conclusion of the first column; catalytic subunit, page 958, first column, first full paragraph), and sequences that do not appear critical for function described (spacer sequences which comprise 60% of the Factor IX sequence, paragraph bridging pages 957 and 958, See, also, page 958, paragraph bridging the first and second columns).

Third, variant Factor IX sequences having function or having reduced function would have been known to those skilled in the art based upon knowledge in the art at the time of the invention. Again, Exhibit B indicates Factor IX amino acids that are not critical for function. For example, a large number of Factor IX amino acid residues are spacer sequences, which comprise 60% of the Factor IX sequence, and are replaceable with most other amino acid residues without resulting in haemophilia B (paragraph bridging pages 957 and 958, See, also, page 958, paragraph bridging the first and second columns). In addition, Exhibit B indicates that the first EGF domain of Factor X can replace that of Factor IX without destroying function (page 957, second column, first full paragraph).

In Exhibit C, the authors report 31 point mutations in Factor IX (see abstract and Table 1). 95 missense mutations are also identified, which occur at evolutionarily conserved amino acids (see abstract). The authors then compared sequences from four and nine species of Factor IX and found that in 40% of the residues, virtually any missense mutation in a minority of residues would cause disease, while in the remaining residues, no missense mutations will cause disease (see, for example, abstract and page 829, second column). As in Exhibit B, 60% of the residues in Factor IX are identified as spacers (see abstract, Exhibit B).

Thus, as corroborated by Exhibits B and C, those skilled in the art would have known which sequences of Factor IX that could be modified without destroying function. Consequently, those skilled in the art would also have known how to produce variant Factor IX sequences having function.

Exhibit B also indicates Factor IX amino acid residues critical for function. In this regard, at least 278 unique Factor IX variant sequences having severely reduced function were known in the art in 1992 (page 960, paragraph bridging the first and second columns). Of these, at least 29 were complete or partial gene deletions, 50 short (less than 20 nucleotides) deletions or insertions, and a large number of single-base missense or non-sense mutations (page 960, second column). As to specific residues critical for function, for example, replacing Glu27 w/Val, disturbs Ca++ binding (page 956, second column); replacing Asp47 w/Gly or Pro55 w/Ala severely reduces factor IX function (page 957, bottom of first column); replacing Cys132 w/Arg or Arg 145 w/Cys or His (page 957, bottom of second column); replacing Pro 287 w/ Leu, Ala 291 w/Pro, or Thr 296 w/Met (page 958, first column, first full paragraph).

Exhibit D lists a database of all known Factor IX point mutations, additions and deletions that cause haemophilia B. A total of 574 patient entries are described and of these 278 are unique molecular events (see abstract).

Thus, in view of the large number of Factor IX mutations and their functional consequences known in the art at the time of the invention, as corroborated by Exhibits B to D, those skilled in the art would have known sequences of Factor IX that could not be modified without destroying or significantly impairing function. Consequently, those skilled in the art would have known the Factor IX variants that would not be desirable for optimal Factor IX function.

Fourth, the cited *Fiers v. Revel* and *Reagents of the Univ. Calif. v. Eli Lilly* case law is also clearly distinguishable form the subject application. For example, the five mammalian Factor IX species, exemplary Factor IX variant disclosed in the subject application, and the Factor IX sequence regions known in the art (as corroborated by Exhibit A) is in dramatic contrast to the number of disclosed species at issue in *Fiers v. Revel.* In *Fiers, Revel* attempted to claim priority to an Israel patent application that did not disclose *any* DNA sequence. In *Lilly* the patent-in-suit disclosed a *single* rat insulin cDNA, which the court held did not provide an adequate written description for generic claims directed to cDNA encoding vertebrate and mammalian insulin. Thus, the subject application, which discloses *five* Factor IX mammalian species and a functional Factor IX variant, as well as Exhibit A, which indicates additional Factor IX sequences were known in the art, is clearly distinguishable from both *Fiers* and *Lily.*

Moreover, there was little if any knowledge in the art regarding the claimed gene/cDNA sequences at issue in both *Fiers* and *Lily.* In stark contrast, knowledge in the art regarding Factor IX sequences as well as structure and function was extensive, as corroborated by Exhibits A to D discussed above.

Thus, in view of the fact that at least five mammalian Factor IX sequences were known in the art at the time of the invention, that the structural and functional domains of mammalian Factor IX was known, and that critical and non-critical Factor IX amino acid sequences and regions were known in the art, as corroborated by Exhibits A to D, those skilled in the art would have known which Factor IX amino acid residues could be modified without destroying function, and also would have known that altering particular amino acid residues would destroy or

substantially impair Factor IX function. Consequently, those skilled in the art would know numerous functional Factor IX sequences for use in the claimed compositions and, as such, would be apprised of the Factor IX genus. Accordingly, an adequate written description for the Factor IX genus is provided and, therefore, Applicants respectfully request that the rejection under 35 U.S.C. §112, first paragraph be withdrawn.

CONCLUSION

In summary, for the reasons set forth herein, Applicants maintain that claims 1 to 9 and 21 to 28 clearly and patentably define the invention, respectfully request that the Examiner reconsider the various grounds set forth in the Office Action, and respectfully request the allowance of the claims which are now pending.

If the Examiner would like to discuss any of the issues raised in the Office Action, Applicant's representative can be reached at (858) 509-4065.

Please charge any additional fees, or make any credits, to Deposit Account No. 03-3975.

Respectfully submitted,

Date: 8-25-03

Robert M. Bedgood, Ph.D.
Reg. No. 43,488
Agent for Applicant

PILLSBURY WINTHROP LLP
11682 El Camino Real
Suite 200
San Diego, CA 92130
(858) 509-4065 Telephone
(858) 509-4010 Facsimile

GENOMICS 6, 133–143 (1990)

Exhibit A

Direct Sequencing of the Activation Peptide and the Catalytic Domain of the Factor IX Gene in Six Species

G. SARKAR, D. D. KOEBERL, AND S. S. SOMMER[1]

Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, Minnesota 55905

Received July 19, 1989; revised September 13, 1989

By means of RNA amplification with transcript sequencing (RAWTS) under low stringency conditions, sequence was obtained directly without cloning for the activation peptide and the catalytic domain of factor IX from six species—sheep, pig, rabbit, guinea pig, rat, and mouse. The data presented demonstrate that, by the appropriate design of oligonucleotides and by performance of a nested PCR under appropriate conditions, it is possible to obtain sequence on a battery of species with a minimum of oligonucleotide primers. A total of 5.2 kb of cross-species sequence was generated with RAWTS. The results indicate that (1) 69% of the amino acids in the catalytic domain, but only 23% of the amino acids in the activation peptide, are *identical* in humans and the six species; (2) the catalytic domain evolves at a slower rate, but the extent and pattern of conservation of amino acids in the activation peptide suggest that the peptide functions as more than a cleavage spacer that separates the heavy and light chains in the catalytically inactive zymogen; (3) 37% of the amino acids in the activation peptide and 34% of the amino acids in the catalytic domain are *factor IX-specific;* i.e., they are either identical or changed in a highly conservative fashion in factor IX, *but not in other related coagulation proteases;* (4) these conserved factor IX-specific amino acids fall into three clusters, which are candidates for involvement in the protein interactions specific to factor IX; (5) there is a human-specific deletion after lysine 142 and a rodent-specific insertion after alanine 161; (6) in guinea pig, the insertion is associated with a seven-amino-acid repeat that corresponds to a perfect repeat of a 21-bp sequence; (7) humans have lost a potential N-glycosylation site that is conserved in the other species; (8) in each species, a few nonconservative changes occur in amino acids that are otherwise completely conserved, suggesting that compensatory mutations may have occurred; and (9) when compared to that of mouse, the amino acid identity with guinea pig factor IX is no greater than that found for the non-rodent species, a result compatible with the postulated increased rate of evolution in rodents.

INTRODUCTION

Human factor IX is an activatable serine coagulation protease that is encoded by a 34-kb gene on the X chromosome (Yoshitake *et al.*, 1985). Factor IX circulates in the plasma as a single-chain, vitamin K-dependent zymogen of 415 amino acids that contains 17% carbohydrate by weight. During clotting, factor IX may be activated by factor XIa in the presence of calcium and by factor VIIa in the presence of calcium and tissue factor (reviewed in Furie and Furie, 1988). Proteolysis by either factor XIa or factor VIIa releases a 35-amino-acid activation peptide. The primary role of activated factor IX in clotting is to activate factor X. The physiological complex includes factor IXa, calcium, phospholipid, and thrombin-activated factor VIII. Factor IXa also binds to anti-thrombin III and it can activate factor VII in the presence of calcium and phospholipid (DiScipio *et al.*, 1978; Masys *et al.*, 1982).

DNA sequence for the coding region of factor IX is available for humans (Kurachi and Davie, 1982; Jaye *et al.*, 1983; Anson *et al.*, 1984; Jagadeeswaran *et al.*, 1984; McGraw *et al.*, 1985) and the amino acid sequence is available for the circulating zymogen from cow (Katayama *et al.*, 1979). If additional sequences were available, it would be possible to better define the conserved amino acids. A subset of these amino acids will be generic in the sense that they are also conserved in factor VII, factor X, and protein C—a group of related coagulation proteases that have the same domains and identical exonic structures (Furie and Furie, 1988). The remainder of the conserved amino acids will be unique

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. M26233, M26234, M26235, M26236, M26237, and M23247.

[1] To whom reprint requests should be addressed.

0888-7543/90 $3.00

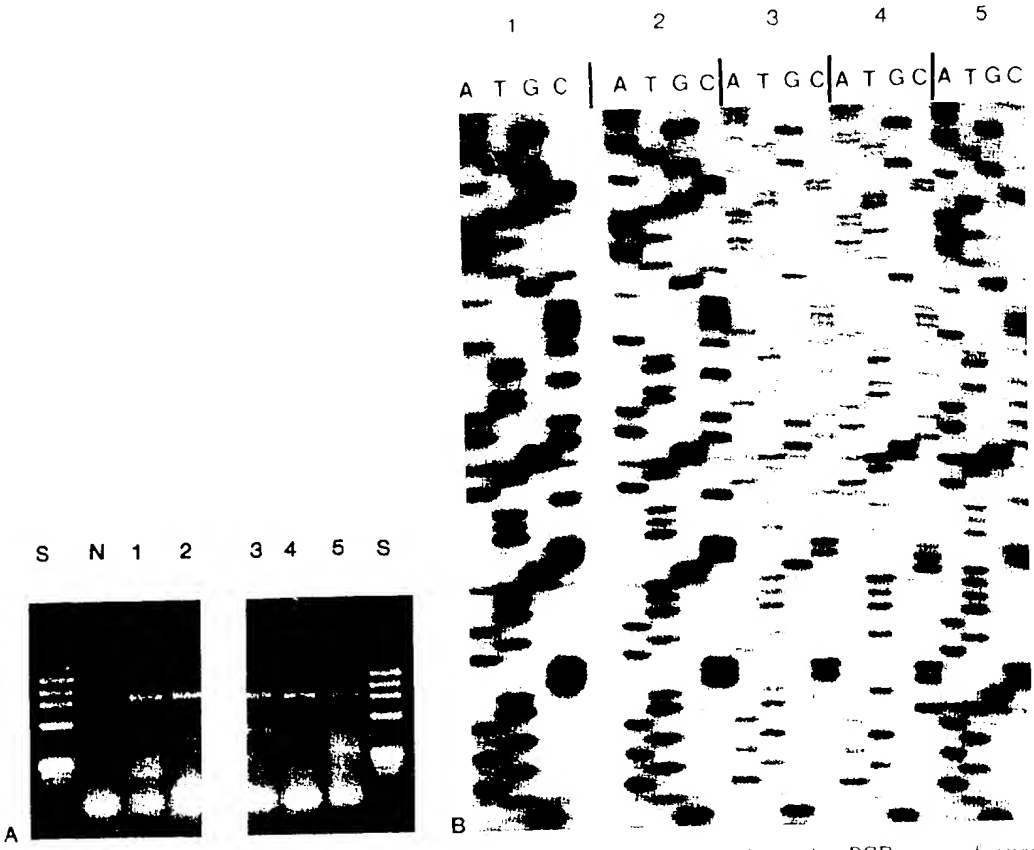


FIG. 1. (**A**) Amplification of the activation peptide and the catalytic domain from multiple species. PCR was performed on liver cDNA with primers to human sequence complementary to the beginning of exon F and the end of the coding region in exon H (OliA and B). If there was no major insertion or deletion, an approximately 900-bp segment migrating just above the third largest size standard was expected. Lane **S**, size standards produced by *Hae*III digestion of ϕX174; lane **N**, control amplification with no template DNA; lane **1**, mouse cDNA; lane **2**, rat cDNA; lane **3**, guinea pig cDNA; lane **4**, rabbit cDNA; and lane **5**, sheep cDNA. (**B**) Sequence of a segment from mouse (**1**), rat (**2**), guinea pig (**3**), rabbit (**4**), and sheep (**5**). Sequence was obtained by RAWTS using the PCR primer (OliB) as the sequencing primer.

to factor IX. These will be candidates for specific interactions of factor IX with factors VII, VIII, X, XI, and anti-thrombin III.

ZooRAWTS is a technique that in principle allows the sequence of homologs in multiple species to be obtained rapidly without the need for cloning (Sarkar and Sommer, 1989). Here we demonstrate more extensively the feasibility of ZooRAWTS by obtaining over 5 kb of DNA sequence from the activation and catalytic domains of the factor IX from six species. The data define amino acids that are specific for factor IX. The catalytic domain of factor IX is found to be highly conserved and the activation domain also has a substantial fraction of conserved amino acids, strongly suggesting that it functions as more than just an inactivating spacer between the heavy and light chains.

MATERIALS AND METHODS

Liver mRNAs were purchased from Clontech. RAWTS was performed as indicated below.

1. First-strand cDNA synthesis: Twenty microliters of 50 μg/ml heat-denatured total RNA or mRNA, 50 m*M* Tris–HCl (pH 8.3), 8 m*M* magnesium chloride, 30 m*M* KCl, 1 m*M* DTT, 2 m*M* each dATP, dCTP, dGTP, dTTP, 50 μg/ml oligo(dT) 12–18, 1000 U/ml RNasin, and 1000 U/ml AMV reverse transcriptase were incubated at 42°C for 1 h, followed by 65°C for 10 min. Subsequently, 30 μl of H_2O was added, generating a final volume of 50 μl.

2. PCR: The above sample (1 μl) was added to 40 μl of 50 m*M* KCl, 10 m*M* Tris–HCl (pH 8.3), 2.0–8.0 m*M* $MgCl_2$ (empirically determined for each set of primers), 0.01% (w/v) gelatin, 200 μ*M* each dNTP, 1 μ*M* each primer (Perkin–Elmer–Cetus protocol). After 10 min at 94°C, 1 U of *Taq* polymerase was added and 30–40 cycles of PCR were performed (denaturation: 1 min at 94°C; annealing: 2 min at 50°C; elongation: 3 min at 72°C) with the Perkin–Elmer–Cetus automated thermal cycler. One primer included a T7 or SP6 promoter

TABLE 1

Mismatches Compatible and Incompatible with Obtaining Sequence

Species and oligonucleotide	Sequence differences[a]	Sequence obtained[b]
I. OliD		
Human	.....T.....	+
Sheep	A.....	+
Pig	.....	+
Rabbit	.G.....T.....	+
Guinea pig	.G...T...T.....	+
Rat	.....C.....	+
Mouse	5' GACTATGAAAATTCTACTGA 3'	+
II. OliF		
Human	.A.....	+
Sheep	5' TGCTGCATTCTGTGGA 3'	+
Pig	.A.....C	−
Rabbit	.AG..G.....	+
Guinea pig	.AA.....	+
Rat	.AG.....	+
Mouse	.AG.....	+
III. OliJ		
Human	.G.....GC.C.	−
Sheep	.....A.....CTC.	−
Pig	.G.....C..CG.C	−
Rabbit	.....CG...	+
Guinea pig	.G...C...C...GA..	−
Rat	.....	+
Mouse	5' ACATAGCTGTTTAGTATTA 3'	+

[a] The oligonucleotide sequence is shown and differences between the oligonucleotide sequence and the species sequence are indicated. OliJ is in the upstream direction so the sequence is complementary to the sequence in Fig. 3.

[b] PCR was performed using the indicated oligonucleotide and either OliA or OliC. Less than 10 pg of the 900-bp T7–SP6 segment was the input DNA. If an amplification product of appropriate size was seen, it was eluted from an agarose gel, transcribed, and sequenced with the PCR oligonucleotide.

as previously described (Stoflet *et al.*, 1988). After the last cycle of PCR, a final 10-min elongation was performed.

3. Transcription: When a nested (internal) sequencing primer was used, 3 μl of the amplified material was added to 17 μl of the RNA transcription mixture. The final mixture contained 40 m*M* Tris-HCl (pH 7.5), 6 m*M* $MgCl_2$, 2 m*M* spermidine, 10 m*M* sodium chloride, 0.5 m*M* of the four ribonucleoside triphosphates, RNasin (1.0 U/μl), 10 m*M* DTT, 10 U of T7 or SP6 RNA polymerase, and diethylpyrocarbonate-treated H_2O. Samples were incubated for 1 h at 37°C and the reaction was stopped by freezing the sample. When the PCR primer was used as a sequencing primer, the PCR-amplified segment of appropriate size was eluted from an agarose gel before the onset of transcription. For segments under 400 bp, the "freeze–squeeze" method was used (Taute and Renz, 1983) and for segments over 400 bp, the GeneClean modification (Bio 101) of the "glass bead" protocol was used. Note that the amplification afforded by transcription obviates the need for quantitative elution during the gel purification.

4. Sequencing protocol [modification of Geliebter (1987)]: The transcription reaction (2 μl) was added to 10 μl of annealing buffer containing the end-labeled reverse transcriptase primer. Annealing and sequencing were performed essentially as described (Stoflet *et al.*, 1988), except that less reverse transcriptase (2.5 units/reaction tube) was used. Note that [γ-^{32}P]ATP is the correct donor for end-labeling.

Nomenclature

Since oligonucleotides accumulate rapidly when GAWTS is used, it is important to have informative names. The following nomenclature readily allows the determination of (i) the size of the amplified fragment, (ii) the appropriateness of any combination of oligonucleotides, and (iii) the origin and direction of the sequence generated. It is of the form G(O)-(I-L)R(C)-SD, where G is gene abbreviation, O is organism, I is

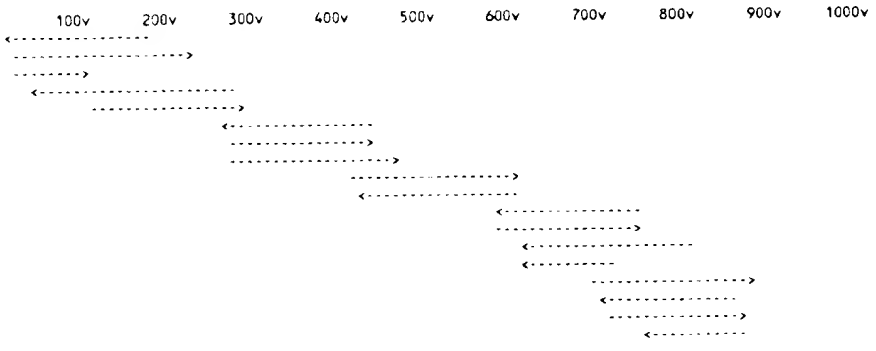


FIG. 2. Sequencing strategy for mouse factor IX. The sequencing strategy for the other species was either identical or there were minor differences. The arrows indicate the direction of the sequencing. Both strands were sequenced for all species.

identifier (one or more) for any noncomplementary 5′ bases, L is length of the noncomplementary bases, R is region of the gene, C is location of the 5′ complementary base, S is total size, and D is 5′ to 3′ direction of the oligonucleotide. The region of the gene (R) is abbreviated by 5′, the region upstream of the gene; by E followed by exon number; by I followed by intron number; or by 3′, the region downstream of the gene. The direction of the oligonucleotide is either U, upstream, or D, downstream. If a transcript has been defined, D is the sense direction and U is the antisense direction. Otherwise, the directions can be arbitrarily defined. Thus F9(Hs)-(T7/TI-37)E6(20365)-51D is an oligonucleotide specific for the factor IX gene of *Homo sapiens,* which has a T7 promoter and a translation initiation sequence of 37 nucleotides (see Sarkar and Sommer, 1989, for sequence). It is complementary to a sequence in exon 6 that begins at base 20365 in the numbering system of Yoshitake *et al.* (1985). The oligonucleotide has a total of 51 nucleotides and it heads downstream relative to factor IX mRNA. As another example, F9(Mm-E8(30949)-19U is an oligonucleotide specific for mouse (*Mus musculus*) factor IX. It is complementary to a sequence in exon 8 of the mouse factor

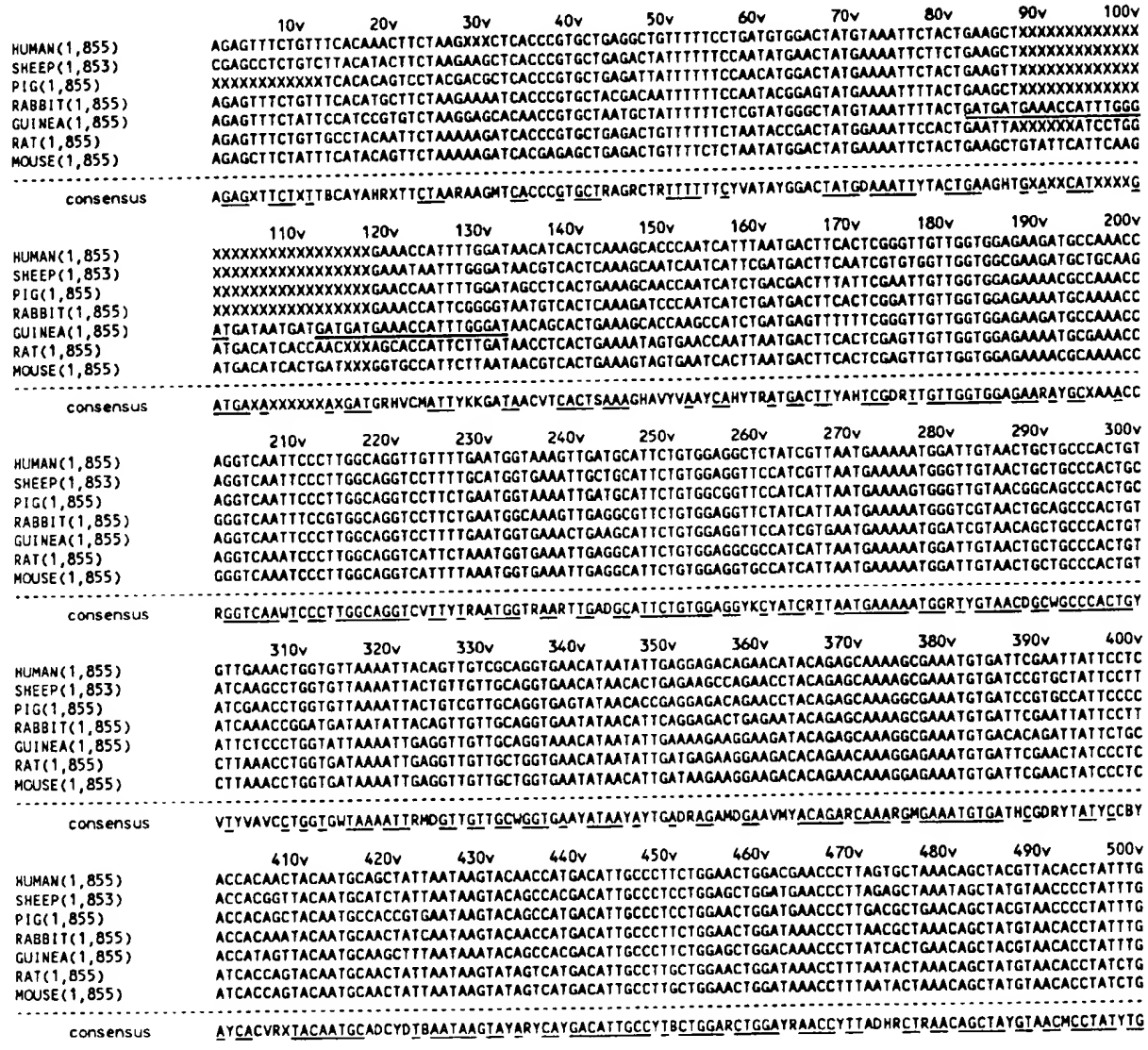


FIG. 3. Alignment of the nucleic acid sequences with human sequences. Nucleotides 40–174 represent the activation peptide. The absence of nucleotides in a species is indicated by X. In guinea pig, the 21-nucleotide in-frame repeat (82–102 and 112–132) that corresponds to the seven-amino-acid repeat is underlined. In the consensus sequence, the underlined nucleotides are identical in all species. The International Union of Biochemistry code is used for the nonidentical nucleotides.

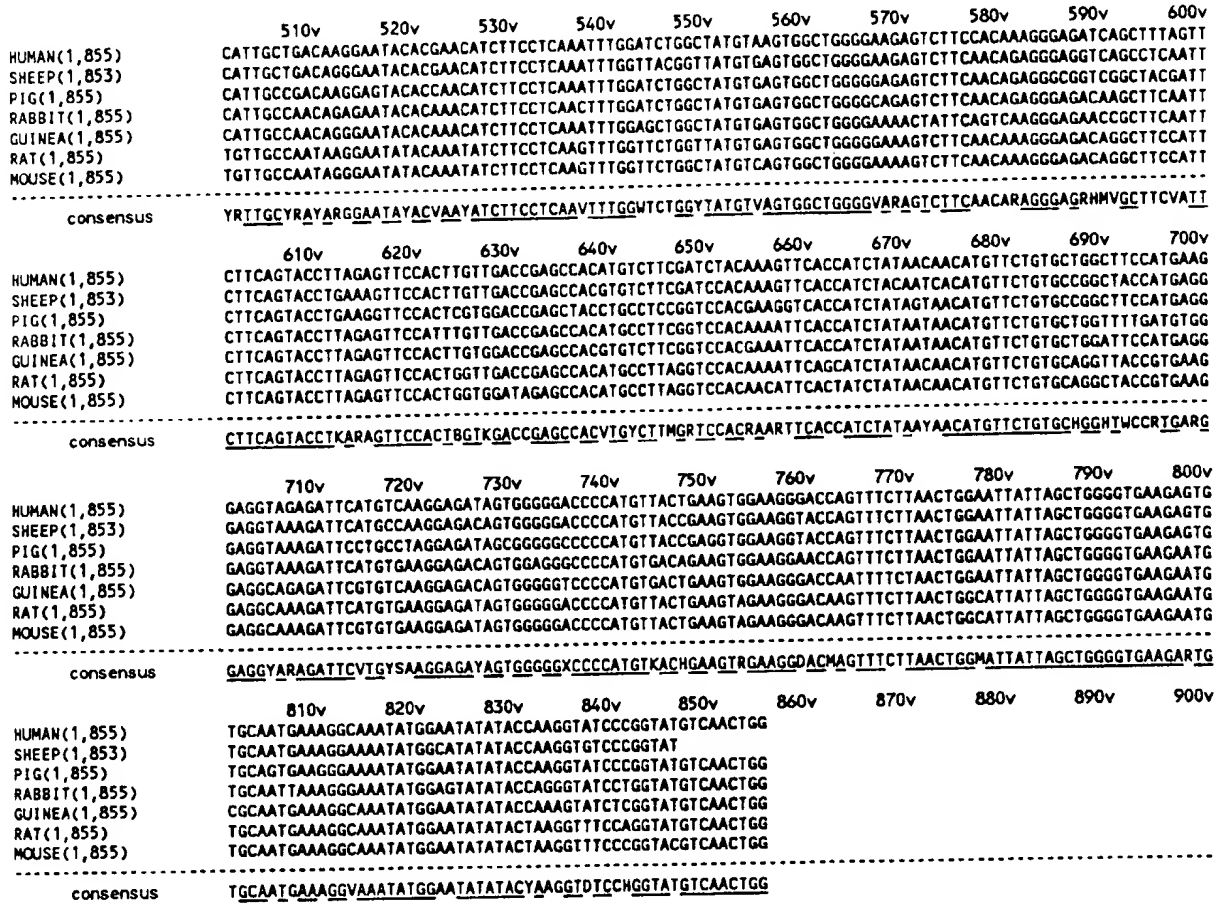


FIG. 3—*Continued*

IX gene that begins at the base that corresponds to the human sequence 30949. The oligonucleotide is 19 nucleotides and the sequence heads upstream relative to *in vivo* transcription.

Oligonucleotides

The oligonucleotides used to sequence both strands (10.4 kb total) are listed with the abbreviation followed by the informative name:

1. OliA: F9(Hs)-(T7/TI-37)E6(20365)-51D—5′ GGATCCTAATACGACTCACTATAGGGAGACC ACCATGCCATTTCCATGTGG3′

2. OliB: F9(Hs)E8(31366)-20U—5′AGTGAGCT TTGTTTTTTCCT3′

3. OliC: F9(Hs)-(Sp6-29)E8(31346)-47U—5′G GTACCATTTAGGTGACACTATAGAATACTAA TCCAGTTGACATACC3′

4. OliD: F9(M)E6(20440)-20D—5′GACTATGAA AATTCTACTGA3′

5. OliE: F9(Hs)E6(20537)-14U—5′TCTTCTCC ACCAAC3′

6. OliF: F9(Sh)-(T7-29)E7(30060)-45D—5′GG ATCCTAATACGACTCACTATAGGGAGATGCT GCATTCTGTGGA3′

7. OliG: F9(Hs)E7(30106)-20U—5′GTTACAAT CCATTTTTCATT3′

8. OliH: F9(M)E8(30849)-16U—5′ATCCAGTT CCAGCAAG3′

9. OliI-1: F9(M)E8(30851)-18U—5′ACAGAACA AAGGAGAAAT3′

10. OliI-2: F9(Hs)E8(30851)-18U—5′ACAGAGC AAAAGCGAAAT3′

11. OliM: F9(Hs)E8(31013)-17D—5′ATCTTCC TCAAATTTGG3′

12. OliN: F9(Hs)E8(31101)-17U—5′CTAAGGT ACTGAAGAAC3′

13. OliO: F9(Hs)E8(31157)-17D—5′AACAACA TGTTCTGTGC3′

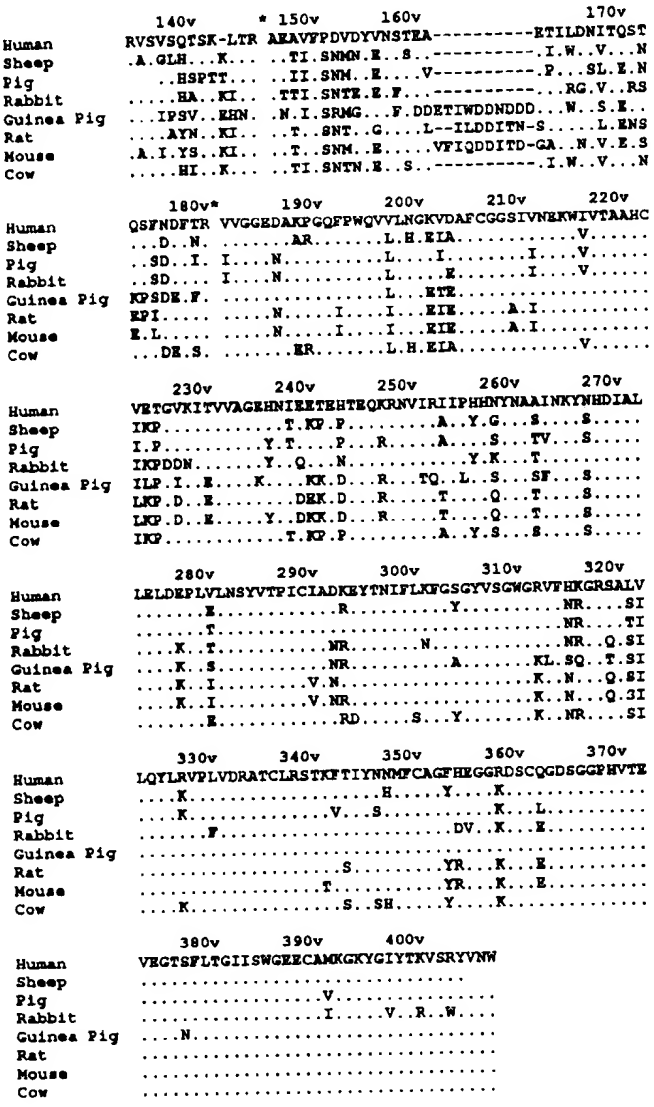


FIG. 4. Alignment of the amino acid sequences with human sequences. The activation peptide is delimited by asterisks. The bovine sequence has previously been published (13). The number system is that of Yoshitake *et al.* (31).

14. OliP: F9(Hs)E8(31215)-16U—5′CTATCTC CTTGACATG3′

15. OliQ: F9(M)E8(31227)-18U—5′CTTCTAC TTCAGTAACAT3′

16. OliR: F9(M)E8(31320)-20U—5′TTAGTATA TATTCCATATTT3′

RESULTS AND DISCUSSION

Sequencing across Species

RNA amplification with transcript sequencing (RAWTS) (Sarkar and Sommer, 1988, 1989) and genomic amplification with transcript sequencing (GAWTS) (Stoflet *et al.*, 1988) are methods of direct sequencing that utilize a phage promoter sequence 5′ to at least one of the PCR primers. RAWTS is a four-step procedure that includes (1) cDNA synthesis with oligo(dT), random primers, or an mRNA-specific oligonucleotide primer; (2) PCR where one or both oligonucleotides contain a phage promoter attached to a sequence complementary to the region to be amplified; (3) transcription with a phage polymerase; and (4) dideoxy sequencing with reverse transcriptase. The procedure for GAWTS is identical except that genomic DNA is the input to step 2. RAWTS/GAWTS has a number of advantages: (1) the transcription step produces an additional level of amplification which obviates the need for purification subsequent to PCR, (2) the amplification afforded by transcription can compensate for a suboptimal PCR, and (3) the generation of a single-stranded template provides more reproducible sequence than that obtained from a double-stranded template. Disadvantages of the technique are the limited number of different sequencing enzymes available and the added expense of attaching phage promoters to the PCR primers. As with all direct sequencing methods, RAWTS is insensitive to the error rate of *Taq* polymerase because the sequence generated is the average for a population of molecules.

RNA amplification with transcript sequencing was performed on mRNA from the livers of pig (*Sus scofa*), sheep (*Ovis aries*), rabbit (*Oryctolagus cuniculus*), guinea pig (*Cavia porcellus*), mouse (*Mus musculus*), and rat (*Rattus norwegicus*). mRNA was chosen as the input nucleic acid rather than genomic DNA because sequence from more than one exon was desired and oligonucleotides complementary to coding sequence were deemed more likely to produce successful amplification across species.

Previously, some initial sequence was obtained across species after a segment, which included sections of exons G and H, was amplified with available human primers (Sarkar and Sommer, 1989). To obtain the sequence from a much larger segment that included the activation peptide and the catalytic domain, the published human and bovine amino acid sequences were compared and PCR primers were made to conserved amino acids at the beginning of exon F (OliA) and the end of the coding region of exon H (OliB). Since precise matches at the 3′ end are critical (Sommer *et al.*, in press), the primers were designed to avoid possible codon redundancy in the 3′ two bases of the oligonucleotide. When the magnesium concentration was varied from 2 to 8 m*M*, segments of the predicted size were seen from each of the species (Fig. 1A). The segment was eluted from an agarose gel, transcribed, and sequenced by reverse transcriptase with OliB (Fig. 1B).

The sequence indicated that the factor IX homologs had been obtained. Since the sequence adjacent to the OliB was well conserved, PCR was performed from cDNA with OliA and a primer with an SP6 promoter sequence OliC. As predicted, factor IX segments (T7-SP6 segments) of about 900 bp (including the phage promoter sequences) from all six species were amplified. Sequence was obtained from one end by transcribing with T7 RNA polymerase and from the other end by transcribing with SP6 polymerase. In some cases, the T7-SP6 segment of a given species could be directly transcribed and sequenced with an internal oligonucleotide that was previously synthesized to delineate mutations in individuals with hemophilia B (Koeberl *et al.*, 1989). Unfortunately, mismatches with the cross-species sequence often precluded efficient priming in the sequencing reaction. PCR was found to be more forgiving of mismatches. Thus, in most cases, sequence was obtained by performing a second round of PCR with an internal oligonucleotide purifying the segment of appropriate size, and sequencing with the internal PCR primer. More specifically, PCR was performed with a given internal oligonucleotide and either the SP6- or the T7-containing oligonucleotide on typically a 3000-fold to 40,000-fold dilution of the 900-bp T7-SP6 segment. The amplified segment of appropriate size was purified from both primer dimers and spurious amplification products by elution from an agarose gel. Either the freeze–squeeze or the glass bead technique was used, depending on the size of the amplified segment (see Materials and Methods). The segment was then transcribed with the appropriate phage polymerase and sequenced with the PCR primer.

TABLE 3

Factor IX Sequences That Are Conserved in Seven Species[a]

	Activation peptide[b] (%)	Catalytic domain[c] (%)
Completely conserved amino acids	23	69
Amino acids with highly conservative changes[d]	17	8
Completely conserved nucleotides	42	67
Completely conserved third base of codon	37	42

[a] The previously published bovine sequence is excluded because only amino acid sequence was determined.
[b] Excluding the insertion in the rodents.
[c] Excluding the last eight amino acids.
[d] The allowed highly conservative substitutions are I/V/L, F/Y, E/D, E/Q, D/N, K/R, T/S, or S/A.

Since the input for the second round of PCR is the T7-SP6 segment, the majority of the DNA molecules contain the segment of interest. If the match between the internal oligonucleotide and the species sequence is good, a 40,000-fold dilution usually provides a strong amplification signal with a minimum of spurious amplification products. Even three or four mismatches can be tolerated if they are not near to the 3′ end where elongation is initiated (Table 1). For mismatches close to the 3′ end, it was necessary to decrease the dilution of the input DNA. With only a 100-fold dilution of the T7-SP6 segment, it was sometimes possible to obtain amplification even if the mismatch occurred at the 3′

TABLE 2

Pairwise Differences in the Sequences for the Activation Peptide and the Catalytic Domain[a]

	Relative to mouse (%)			Relative to humans (%)		
	Amino acid	Nucleic acid	Third base of codon	Amino acid	Nucleic acid	Third base of codon
Mouse	—	—	—	37/20	19/15	17/25
Rat	29/2	15/3	17/6	40/20	25/13	23/21
Guinea pig	60/14	33/16	31/30	15/18	22/13	14/22
Rabbit	41/15	26/15	29/28	46/20	21/12	14/20
Pig	37/15	24/19	26/34	43/18	21/14	20/28
Sheep	37/18	22/19	23/32	40/16	17/13	14/19
Human	37/20	19/15	14/25	—	—	—
Average[b]	42/16	25/17	25/30	43/19	21/13	17/23

[a] The comparisons for the activation peptide are based on only the 35 amino acids that all these peptides have in common. For each category, the percentage of different amino acids or nucleotides for the activation peptide is presented to the left of the shill followed by the corresponding value for the catalytic domain. The bovine sequence (13) was excluded from these comparisons because only amino acid sequence is available.
[b] The average for mouse excludes the closely related rat sequence.

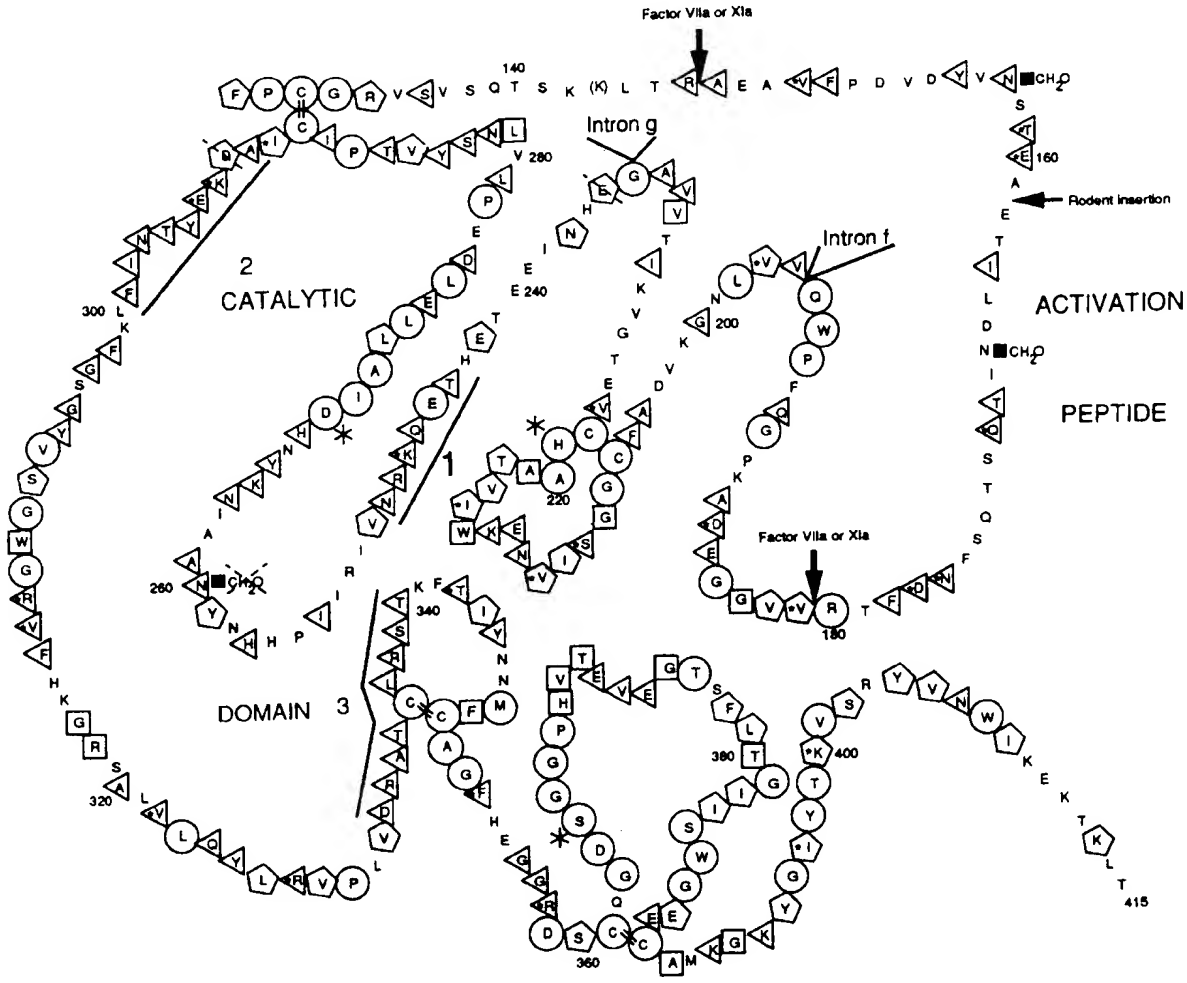


FIG. 5. Types of amino acid conservation in factor IX. The bovine sequence (13) is included in this analysis for a total of eight species. The sequence of human factor IX is shown. The geometric shapes indicate the type of amino acid conservation (see text for description). Those in triangles are conserved in factor IX but not in factor VII, factor X, and protein C. Three clusters of factor IX-specific amino acids are underlined. Asterisks inside the geometric shape indicate that highly conservative changes occur in the factor IX sequences from the eight species. The conserved catalytic triad is indicated by a large asterisks. The locations of the glycosylation sites are indicated by CH_2O. The cleavage sites for factors VII and XI and the location of introns F and G are indicated. The lysine after amino acid 142 that is deleted in humans is shown in parentheses. The point of the rodent insertion is shown and a conserved glycosylation site that is lost in humans is indicated by CH_2O with an X through it. The disulfide bonds are indicated by C=C. The three disulfide loops are extremely well conserved. All have coagulation protease generic amino acids and all have at least three consecutive factor IX-specific conserved amino acids.

end of the oligonucleotide (Sarkar *et al.*, manuscript in preparation).

Sequence Comparisons

By sequencing of transcripts generated by T7 and SP6 RNA polymerase, respectively, data were obtained from both strands (Fig. 2). The nucleic acid and the deduced amino acid sequences were aligned with the human sequence (Figs. 3 and 4). Pairwise comparisons of the sequences with mouse factor IX sequence indicate that rat factor IX is more closely related than the others (Table 2), whereas guinea pig factor IX sequence has diverged as much as the rabbit, ungulate, and primate sequences. This is compatible with a postulated increased rate of evolution in rodents (Li *et al.*, 1987; Wu and Li, 1985). An alternate, perhaps less likely possibility, is that there are major misinterpretations of the rodent fossil record (Wilson *et al.*, 1987).

Pairwise comparisons with human factor IX sequence show the same average amino acid divergence as that seen when mouse was the comparison standard (Table 2). Curiously, there is less deviation at the third base of the codon when humans are the comparison standard, but the average divergence of the first and second bases is similar in comparisons with human or

TABLE 4

Sequence of Potential N-Glycosylation Sites

	Position of asparagine[a]						
Species	157	In[b]	167	172	228	249	260
Human	NST		NIT				
Rabbit	NFT		NVT		NIT		NAT
Guinea pig	NFT		NST			NVT	NAS
Rat	NST	NST	NLT				NAT
Mouse	NST		NVT				NAT
Pig	NST		—	NQS			NAT
Sheep	NSS		NVT	NQS			NAS
Cow	NSS		NVT	NQS			NAS

[a] The numbering system is that of human factor IX (31).
[b] This asparagine is located in the rodent-specific insertion.

mouse. The difference in the third base of the codon may reflect differences in the codon biases found in mouse (data not shown).

Patterns of Conservation

For the activation peptide, 42% of the bases are invariant in all seven species listed in Table 2 while the catalytic domain is more conserved, as 67% of the bases are invariant ($P < 0.001$ by the binomial distribution) (Table 3). The deduced amino acid sequence indicates that there is a deletion in humans of a lysine located three amino acids before the beginning of the activation peptide (Fig. 5). In addition, an insertion is seen in rodents. The size and the sequence of the insertion are not well conserved. In the guinea pig, most of the insertion is part of a seven-amino-acid repeat which is presumably of recent origin because the corresponding cDNA sequence is a precise 21-base repeat (Fig. 3). These were the only deletions and insertions found. The presence of a nine-amino-acid insertion and the marked conservation of the carboxy terminus relative to humans agree with an abstract reporting the cloning of mouse factor IX cDNA (Wu *et al.*, 1988).

Twenty-three percent of the activation peptide amino acids and 69% of the catalytic domain amino acids were identical in the seven species (Table 3). In the activation peptide, another 17% of the amino acids exhibited only highly conservative substitutions while only 8% of those in the catalytic domain exhibited such highly conservation changes, i.e., a ratio of identity to high conservative changes of 8.6:1. The remaining 60 and 23% of the amino acids in the activation peptide and catalytic domain were either less conserved or unconserved. While the catalytic domain is more conserved, the magnitude of conservation in the activation peptide and the absence of clustering near the cleavage sites suggest that it may function as more than a spacer between the heavy and light chain. In *in vitro* activation of bovine factor IX with factor XIa, it was observed that the cleaved activation peptide was noncovalently associated with factor IXa (Amphlett *et al.*, 1979). If noncovalent association after cleavage were of physiological importance, such an interaction might possibly account for some of the conservation that is seen.

The two known N-glycosylation sites in the human activation peptide (Balland *et al.*, 1988) are conserved in all species with the exception of pig where the second site is lost or perhaps displaced six amino acids downstream (Table 4). In humans, a conserved potential glycosylation site at asparagine 260 (catalytic domain) is lost. This site is known to be glycosylated in cow (Mizuochi *et al.*, 1983), strongly suggesting that a common glycosylation site found in mammals has been lost in human factor IX. While all the sequenced factor IX proteins are likely to be glycosylated, the precise location of the glycosylation seems not to be crucial, save perhaps the site at asparagine 157. The N-glycosylation sites constrain the evolution of the activation peptide. The conservation of glycosylation may reflect its importance in maintaining an appropriate turnover rate for factor IX in the circulation.

In each species, there are a few divergences from otherwise completely conserved amino acids. In humans, proline is present instead of serine at position 151. Likewise, threonine is present in place of proline 225 and valine is present in place of isoleucine 322. Such substitutions may be tolerated in humans because an apparently compensatory change occurred at another position in factor IX (perhaps the serine to proline and the proline to threonine substitutions constitute a pair of compensating mutations) or, conceivably, in one of the other proteins that interacts with factor IX. This hypothesis predicts, for example, that an *in vitro* mutation of proline 151 to serine would inactivate human factor IX. The simultaneous mutation of proline 151 and a compensatory change would be required for human factor IX to retain its activity.

Types of Amino Acid Conservation

An alignment was performed for the amino acid sequences of human factors VII, IX, X, protein C, and trypsin (Hagen *et al.*, 1986; Fung *et al.*, 1985; Beckman *et al.*, 1985; Emi *et al.*, 1986) and bovine factors IX, X, protein C, and trypsin (Katayama *et al.*, 1979; Fung *et al.*, 1985; Long *et al.*, 1984; Mikes *et al.*, 1966, as modified by Kossiakoff *et al.*, 1977). Thus a total of seven coagulation proteases and two trypsins were aligned. Conserved amino acids were identified in the catalytic domain, but not in the activation peptide (data not shown). Different levels of amino acid conservation were seen in the catalytic domain (Fig. 5). Those that

were identical or highly conserved can be viewed as generic. Twenty percent of the amino acids in the catalytic domain are identical in the four coagulation proteases (Fig. 5, circles). A comparison with Fig. 5 indicates that all these amino acids were identical in the eight species of factor IX. Sixteen percent of the amino acids vary in a highly conservative fashion in the coagulation proteases (Fig. 5, pentagons). Of these, some show the same variation in factor IX [e.g., amino acids 181 (V or I) and 400 (K or R)] (pentagons with asterisks inside) while the others were identical in all species of factor IX [e.g., amino acid 210 (I but not L) and amino acid 218 (T but not S)] (pentagons without asterisks). There were two exceptions to the conservation. Both glutamate and aspartate appear in the five coagulation proteases at the position that aligns with the factor IX amino acids 235 and 292. However, guinea pig has a lysine in position 235 and four species have an asparagine in position 292.

Seven percent of the amino acids are identical in the four coagulation proteases except for a nonconservative change in *one* of the following: factor VII, factor X, and protein C (squares). All of these amino acids were identical in the eight factor IX sequences. In total, 43% of the amino acids of the catalytic domain are generic for the coagulation proteases. In addition, 34% of the conserved amino acids are specific for factor IX. Ninety percent of these factor IX-specific sites are identical (triangle) in all eight species, while 10% undergo highly conservative substitution (triangles with asterisks). In the activation peptide, 3% of the amino acids are generic and 37% are specific for factor IX. Thus, both domains have about the same frequency of factor IX-specific amino acids, while the frequency of coagulation protease generic amino acids varies markedly.

From the factor IX-specific sequences, clusters of conserved amino acids were arbitrarily defined as requiring at least four consecutive conserved amino acids for nucleation and at least two consecutive nonconserved or protease generic amino acids for termination (Fig. 5, segments 1–3). *In vitro* mutagenesis or analysis of CRM+ hemophiliac factor IX proteins with mutations at these clusters will allow an assessment of the importance of these clusters in the specific interactions of factor IX with factors VII, VIII, X, XI, and antithrombin III.

In conclusion, the remarkable tolerance demonstrated herein for mismatches implies that sequence can be generated for a battery of species with a minimum of oligonucleotides. mRNA was the source of the sequence in these experiments, but cDNA or genomic libraries or genomic DNA could also be used. The enhanced rate of cross-species sequencing afforded by ZooRAWTS will be applicable generally to studies concerned with structure–function and/or evolution.

ACKNOWLEDGMENTS

This work was aided in part by MOD5-647 from the March of Dimes Birth Defects Foundation. D.D.K. has a predoctoral fellowship (18-3) from the March of Dimes.

REFERENCES

1. AMPHLETT, G. W., BYRNE, R., AND CASTELLINO, F. J. (1979). The binding of calcium to the activation products of bovine factor IX. *J. Biol. Chem.* **254:** 6333–6336.
2. ANSON, D. S., CHOO, K. H., REES, D. J. G., GIANNELLI, F., GOULD, J. A., HUDDLESTON, J. A., AND BROWNLEE, G. G. (1984). The gene structure of human anti-haemophiliac factor X. *EMBO J.* **3:** 1053–1060.
3. BALLAND, H., FAURE, T., CARVALLO, D., CORDIER, P., ULRICH, P., FOURNET, B., DELASALLE, H., AND LECOCQ, J-P. (1988). Characterisation of two differently processed forms of human recombinant factor IX synthesised in CHO cells transformed with a polycistronic vector. *J. Biochem.* **172:** 565–572.
4. BECKMANN, R. J., SCHMIDT, R. J., SANTERRE, R. F., PLUTZY, J., CRABTREE, G. R., AND LONG, G. L. (1985). The structure and evolution of a 461 amino acid human protein C precursor and its messenger RNA, based upon the DNA sequence of cloned human liver cDNAs. *Nucleic Acids Res.* **13:** 5233–5247.
5. DISCIPIO, R. G., KURACHI, K., AND DAVIE, E. W. (1978). Activation of human factor IX (Christmas factor). *J. Clin. Invest.* **61:** 1528–1538.
6. EMI, M., NAKAMURA, Y., OGAWA, M., YAMAMOTO, T., NISHIDE, T., MORI, T., AND MATSUBARA, K. (1986). Cloning, characterization and nucleotide sequences of two cDNAs encoding human pancreatic trypsionogens. *Gene* **41:** 305–310.
7. FUNG, M. R., HAY, C. W., AND MACGILLIVRAY, R. T. A. (1985). Characterization of an almost full-length cDNA coding for human blood coagulation factor X. *Proc. Natl. Acad. Sci. USA* **82:** 3591–3595.
8. FURIE, B., AND FURIE, B. C. (1988). The molecular basis of blood coagulation. *Cell* **53:** 505–518.
9. GELIEBTER, J. (1987). Dideoxynucleotide sequencing of RNA and uncloned cDNA. *Focus* **9**(1): 5–8.
10. HAGEN, F. S., GRAY, C. L., O'HARA, P., GRANT, F. J., SAARI, G. C., WOODBURY, R. G., HART, C. E., INSLEY, M., KISIEL, W., KURACHI, K., AND DAVIE, E. W. (1986). Characterization of a cDNA coding for human factor VIII. *Proc. Natl. Acad. Sci. USA* **83:** 2412–2416.
11. JAGADEESWARAN, P., LAVELLE, D. E., KAUL, R., MOHANDAS, T., AND WARREN, S. T. (1984). Isolation and characterization of human factor IX cDNA: Identification of *Taq*I polymorphism and regional assignment. *Somat. Cell Mol. Genet.* **10:** 465–473.
12. JAYE, M., DELASALLE, H., SCHAMBER, F., BALLAND, A., KOHLI, V., FINDELL, A., TOLSTOSHEV, P., AND LECOCQ, J-P. (1983). Isolation of a human anti-haemophiliac factor IX cDNA clone using a unique 52-base synthetic oligonucleotide probe deduced from the amino acid sequence of bovine factor IX. *Nucleic Acids Res.* **11:** 2325–2335.
13. KATAYAMA, K., ERICSSON, L. H., ENFIELD, D. L., WALSH, K. A., NEURATH, H., DAVIE, E. W., AND TITANI, K. (1979). Comparison of amino acid sequence of bovine coagulation factor IX (Christmas Factor) with that of other vitamin K-dependent plasma proteins. *Proc. Natl. Acad. Sci. USA* **76:** 4990–4994.
14. KOEBERL, D. D., BOTTEMA, C. D. K., BUERSTEDDE, J-M., AND SOMMER, S. S. (1989). Functionally important regions of the

factor IX gene have a low rate of polymorphism and a high rate of mutation in the dinucleotide CpG. *Amer. J. Hum. Genet.* **45:** 448–457.

15. KOSSIAKOFF, A. A., CHAMBERS, J. L., KAY, L. M., AND STROUD, R. M. (1977). Structure of bovine trypsinogen at 1.9 A resolution. *Biochemistry* **16:** 654–664.
16. KURACHI, K., AND DAVIE, E. W. (1982). Isolation and characterization of cDNA coding for human factor IX. *Proc. Natl. Acad. Sci. USA* **79:** 6461–6464.
17. LI, W-H., TANIMURA, M., AND SHARP, P. M. (1987). An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25:** 330–342.
18. LONG, G. L., BELAGAJE, R. M., AND MACGILLIVRAY, R. T. A. (1984). Cloning and sequencing of liver cDNA coding for bovine protein C. *Proc. Natl. Acad. Sci. USA* **81:** 5653–5656.
19. MASYS, D. R., BAJAJ, S. P., AND RAPAPORT, S. I. (1982). Activation of human factor VII by activated factors IX and X. *Blood* **60:** 1143–1150.
20. MCGRAW, R. A., DAVIS, L. M., NOYES, C. M., LUNDBLAD, R. L., ROBERTS, H. R., GRAHAM, J. B., AND STAFFORD, D. W. (1985). Evidence for a prevalent dimorphism in the activation peptide of human coagulation factor IX. *Proc. Natl. Acad. Sci. USA* **82:** 2847–2851.
21. MIKES, O., HOLEYSOVSKY, V., TOMASEK, V., AND SORM, F. (1966). Covalent structure of bovine trypsinogen: The position of the remaining amides. *Biochem. Biophys. Res. Commun.* **24:** 346–352.
22. MIZUOCHI, T., TANIGUCHI, T., FUJIKAWA, K., TITANI, K., AND KOBATA, A. (1983). The structures of the carbohydrate moieties of bovine blood coagulation factor IX (Christmas factor). *J. Biol. Chem.* **258:** 6020–6024.
23. SARKAR, G., AND SOMMER, S. S. (1988). RNA amplification with transcript sequencing (RAWTS). *Nucleic Acids Res.* **16:** 5197.
24. SARKAR, G., AND SOMMER, S. S. (1989). Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science* **244:** 331–334.
25. SOMMER, S. S., CASSADY, J., SOBELL, J. L., AND BOTTEMA, C. D. K. A novel method of detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. *Mayo Clinic Proc.*, in press.
26. STOFLET, E. S., KOEBERL, D. D., SARKAR, G., AND SOMMER, S. S. (1988). Genomic amplification with transcript sequencing. *Science* **239:** 491–494.
27. TAUTZ, D., AND RENZ, M. (1983). An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. *Anal. Biochem.* **132:** 14–19.
28. WILSON, A. C., OCHMAN, H., AND PRAGER, E. M. (1987). Molecular time scale for evolution. *Trends Genet.* **3:** 241–247.
29. WU, C-I., AND LI, W-H. (1985). Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* **82:** 1741–1745.
30. WU, S. M., WARE, J., AND STAFFORD, D. (1988). Sequence of mouse factor IX cDNA. *Blood (Suppl)* **72:** 314a.
31. YOSHITAKE, S., SCHACH, B. G., FOSTER, D. C., DAVIE, E. W., AND KURACHI, K. (1985). Nucleotide sequence of the gene for human factor IX (anti-hemophiliac factor B). *Biochemistry* **24:** 3736–3750.

Blood Coagulation and Fibrinolysis, 4, 953 – 974

Review Paper

Exhibit B

Biology of factor IX

K. Kurachi, S. Kurachi, M. Furukawa and S-N. Yao

(Received 22 February 1993; accepted in revised form 5 October 1993)

Within the past 20 years or so, factor IX has been at the centre of particularly intensive studies of its physiology, pathology and biochemistry as well as its molecular genetics and biology. With the complete nucleotide sequence of its human gene determined in 1985 and the molecular defects of over 600 abnormal human factor IX genes analysed to date, factor IX is among the few mammalian proteins which have been exhaustively studied in almost every aspect. The enormous amount of information we now have on this medium-sized plasma protein sheds light on how a gene and its protein evolve, how the protein carries out a highly regulated, specific and pivotal role in the delicately balanced blood coagulation reaction, and the correlation between clinical presentations and its highly diverse molecular mechanism of defects. This wealth of knowledge makes factor IX an excellent model for deeper study, such as truly quantitative analysis of its structure–function relationship and *in vivo* function and regulation. It will also provide a sound foundation which may lead to improved treatment of haemophilia B and perhaps to its cure. This paper attempts to review the recent progress in research on factor IX.

Key words: Haemophilia, factor IX, molecular genetics, review.

Introduction

Blood coagulation is the principal mechanism which follows the initial platelet plug formation to stop blood loss after vascular injury.[1] The basic mechanism of blood coagulation and its regulation involves more than 20 protein factors in addition to calcium ion and phospholipids.[1,2] In this mechanism, factor IX plays a crucial role occupying a key juncture of the intrinsic pathway involving factor XI and the extrinsic pathway involving factor VII and tissue factor (Figure 1). After activation by these pathways, factor IX in turn activates factor X in the presence of factor VIII, Ca^{2+} and phospholipid surface. A deficiency of factor IX in the circulation results in a bleeding disorder, haemophilia B.

Recently, factor XI was shown to be activated by thrombin, resulting in significant revisions in the coagulation cascade (Figure 1).[3,4] These revisions have indicated the important roles of thrombin and factor XI in the initiation and maintenance of blood coagulation, and have provided an explanation for a lack of bleeding disorders due to deficiencies of factor XII, prekallikrein, and high molecular weight kininogen. Because both haemophilia A (deficiency of factor VIII) and haemophilia B patients bleed in spite of the normal amount of factor VII in their circulation and sufficient amount of tissue factor available, generation of factor Xa catalysed by the pathway involving factor IXa–factor VIII complex is obviously crucial for the stable maintenance of coagulation. Activation of factor IX and/or factor X by the factor VIIa–tissue factor complex upon vascular injury may be essential for the initiation of coagulation by generating the initial minute amount of thrombin which may, in turn, activate factor XI to factor XIa leading to subsequent production of factor IXa.[2] The activation pathway of factor X by factor VIIa–tissue factor, however, appears to be tightly controlled by lipoprotein-associated coagulation inhibitor (also called extrinsic pathway inhibitor), suggesting its transient role, if any, in maintenance of coagulation.[2,5,6] The activation pathway of factor IX by the factor VIIa–tissue factor complex, however, may still play an important role in maintenance of coagulation *in vivo*.[7]

This article reviews current knowledge of the biology of factor IX: its structure–function relationships, gene structure and abnormal genes, regulation of

The authors are with the Department of Human Genetics, University of Michigan Medical Center, 3712 Medical Science Building II, Ann Arbor, MI 48109–0618, USA. Tel: (+1) 313–747–3153; Fax: (+1) 313–763–3784. Address correspondence to Professor K. Kurachi.

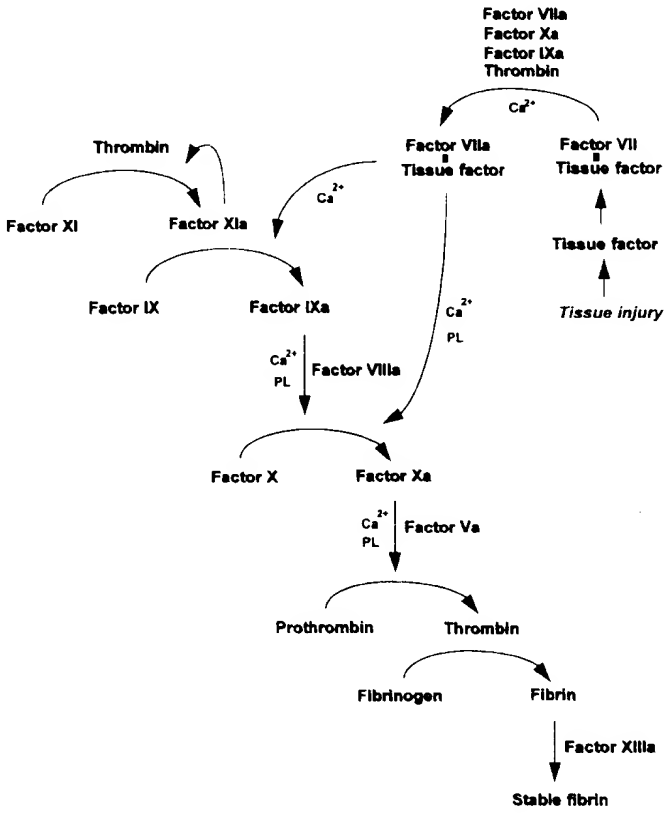


Figure 1. Basic mechanism of blood coagulation. PL indicates phospholipids. Activated forms of coagulation factors are shown with a suffix 'a'.

the gene and the current status of developing new therapy for haemophilia B.

Structure of factor IX

Human factor IX is synthesized as a prepro form of a single polypeptide chain (Figure 2).[8,9] Prepro factor IX is composed of several distinct domain (or module) structures. These include preleader (also called signal peptide) which spans amino acid (aa) −46 through −19, proleader (also called propeptide) spanning aa −18 through −1, Gla domain (the amino-terminal region of about 40 amino acid residues starting at aa +1) containing twelve γ-carboxylated glutamic acid (Gla) residues, a short hydrophobic sequence, two epidermal growth factor-like domains (each about 40 amino acid residues in length), a linking sequence, activation peptide (35 amino acid residues in length) and catalytic subunit of 235 amino acid residues. Intron positions relative to the amino acid sequences divide these domain structures in a characteristic manner. During secretion, both signal peptide and propeptide are cleaved off and the mature factor IX (plasma form of 415 amino acid residues in length) is produced. Although factor IX cDNA has three Met codons in the same reading frame clustered in the amino-terminal end region at aa −46, −41 and −39, the third Met residue at nucleotide (nt) −39 has a sequence (ATCATGG) which matches best with the Kozak consensus sequence (NNPuNNATGGNN),[10] suggesting that this Met could be the primary translation start site. This is further supported by the sequences of different species which have conserved Met at aa −39 but are missing a Met residue at −41 (dog, rat and mouse) and at −46 (dog and rat).[11] According to the ATG scanning model,[12] the first ATG at −46 may still be used for translation initiation albeit at low level. Prepro factor IX undergoes several co- and post-translational modifications. Its signal peptide is cleaved off by signal peptidase during the secretion of the nascent polypeptide chain, and the propeptide sequence is eventually cleaved off by a processing protease during the secretion of factor IX protein. Proteases which may be responsible for removing propeptides with a dibasic amino acid sequence at its carboxyl-terminus have been isolated and their cDNAs cloned.[13] These proteases are either metalloendopeptidase[13] or subtilisin type proteases (PACE).[14] In co-expression experiments with factor IX in Chinese hamster ovary (CHO) cells, PACE could enhance propeptide cleavage which takes place late in the secretory pathway.[15] Whether both or only one of these different endopeptidases are responsible for processing in the liver is not known.

The mature plasma factor IX is a single polypeptide chain starting with Tyr at aa +1 and ending with Thr at aa +415. Interestingly, all other homologous vitamin K dependent proteins of the coagulation system have Ala at this position (Figure 2).[9] Study by mutagenesis has shown that replacement of the Tyr residue at this position with Ala significantly improves the cleavage of the propeptide with little effect on γ-carboxylation of recombinant factor IX.[16] Why factor IX maintains a Tyr residue at this position is not known. It may have some biological significance in the overall regulation of coagulation. The plasma factor IX is secreted from hepatocytes into the bloodstream. During blood coagulation, the plasma factor IX undergoes limited proteolyses which free a 35 amino acid residue-long activation peptide, converting itself to the activated form, factor IXa. This is catalysed by either factor XIa in the presence of Ca^{2+} ions or factor VIIa in the presence of tissue factor and Ca^{2+} ions. Factor IXa is composed of a light chain (the amino-terminal half, aa 1–145) containing five structural domains with various functions to regulate the factor IX and a heavy chain (the carboxyl-terminal half, aa 180–415) containing the protease domain (catalytic subunit).

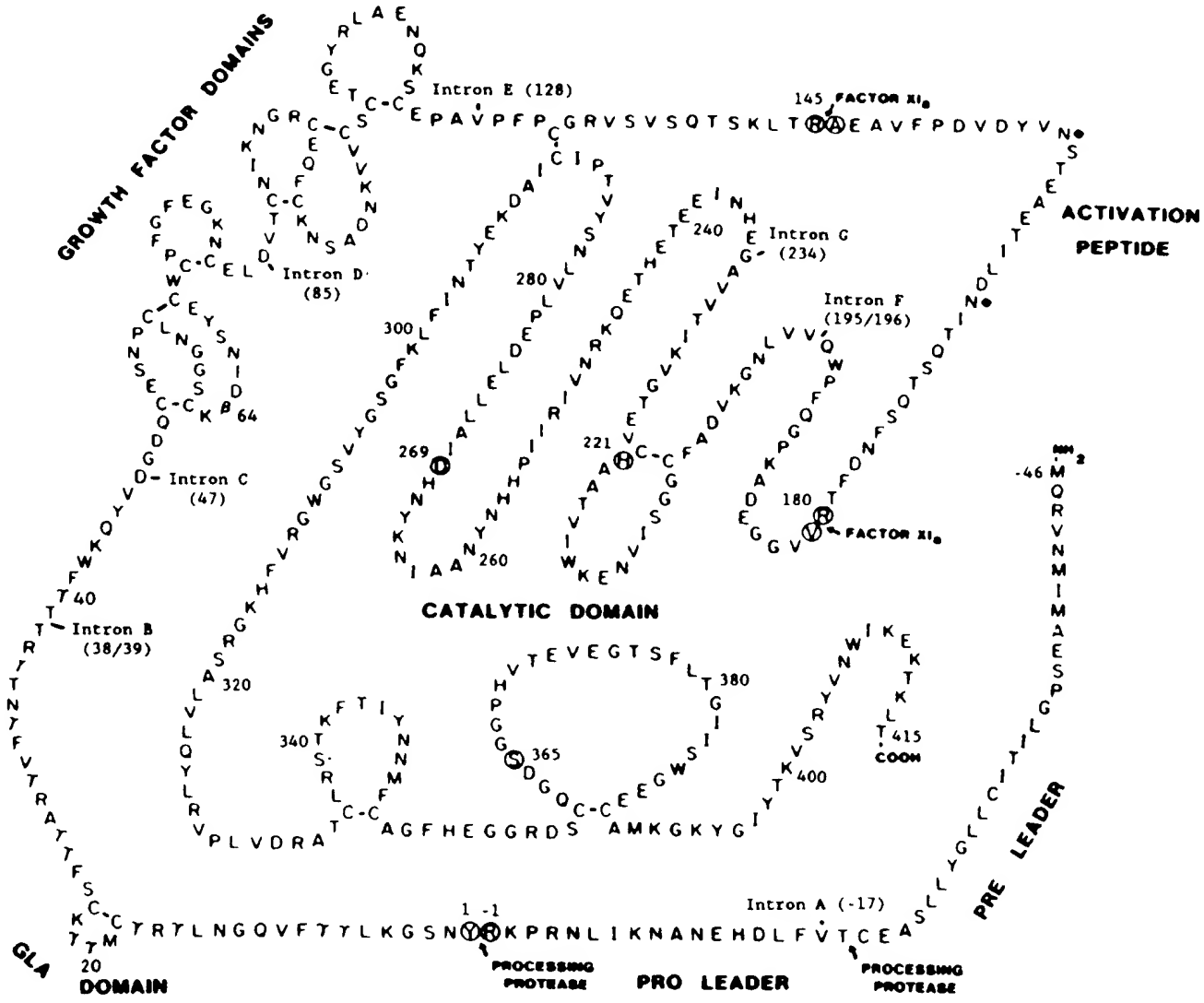


Figure 2. Amino acid sequence and tentative domain structure of human prepro factor IX. Numbering of amino acid sequence: positive numbers for the mature form of factor IX starting with Tyr at +1; negative numbers in reverse direction for preproleader sequence starting with Arg at −1. Intron positions in relation to the amino acid sequence are shown with residue numbers in parentheses. Arrows show the locations of peptide bonds which are cleaved during processing and activation of the prepro factor IX. Circled residues in the catalytic subunit indicate key amino acid residues involved in the active site. Modified from Yoshitake *et al.*[9]

The propeptide sequence of factor IX (aa −18 to −1) plays an important role in the vitamin K dependent γ-carboxylation of twelve glutamic acid residues contained in the Gla domain.[17,18] The importance of propeptide is also shown for other similar vitamin K dependent factors such as protein C.[19] Gla residues in the Gla domain play an essential role in the biological function of factor IX as Ca^{2+} binding sites. Studies on abnormal factor IX genes and a series of mutagenesis analyses carried out on the Gla domain and propeptide sequence have shed light on the mechanism responsible for the modification.[17,18] The propeptide alone without the Gla domain can serve as the recognition site for vitamin K dependent carboxylase which is embedded in the rough endoplasmic reticulum membrane. This is supported by the finding that synthetic propeptides alone can augment the carboxylase activity.[17,18] The maintenance of the approximate size of the intact propeptide is apparently important for its function, and some residues of the propeptide including those at aa −18, −17, −16 (Phe; conserved among vitamin K dependent proteins), −15, and −10 (Ala; conserved) are critical for the reaction. In these studies, Arg residues at aa −4 and −1 have been shown not to be critical

for the γ-carboxylation reaction. Interestingly, however, mutant factor IX molecules, factor $IX_{Cambridge}$ (Arg-1 changed to Ser)[20] and factor $IX_{San\ Dimas}$ (Arg-4 changed to Glu)[21] which are transported into the circulation with propeptides uncleaved, show decreased levels of γ-carboxylation. Mutations at aa −1 and −4 not only inhibit the proper cleavage of the peptide bond between Arg-1 and Tyr+1, but also may affect γ-carboxylation to some extent by changing the conformation of the propeptide and Gla domains. In γ-carboxylation, propeptide bound to carboxylase may function to anchor the nascent factor IX polypeptic chain so that the active site of carboxylase specifically recognizes the unmodified Gla domain region and scans the domain modifying the twelve Glu residues in the region to Gla residues.[17,18]

Recently, a cDNA clone for membrane-integrated, vitamin K dependent γ-glutamyl carboxylase was isolated.[22] When expressed in COS-1 cells and CHO cells, this carboxylase can augment the *in vitro* γ-carboxylase activity of microsomal preparations by 17- and 16-fold, respectively,[23] agreeing reasonably well with the previous observation.[22] In contrast, transient co-transfection of the γ-carboxylase expression vector into factor IX-expressing CHO cells did not improve the specific procoagulant activity of secreted factor IX, suggesting that the γ-carboxylation of factor IX is not limited by the expression of the vitamin K dependent γ-carboxylation alone.[23]

A very high expression of factor IX in hepatoma cells (> 100 μg/ml medium) in culture results in a poor specific activity (only 1.5%) of factor IX, apparently due to poor γ-carboxylation.[24] When a relatively high level of recombinant factor IX (1–3 μg/10^6 cells/day) is produced by various heterologous cell lines such as BHK cells and CHO cells, its specific activity is also low, varying in a range of 25–70%.[25] Furthermore, a significant fraction (20%) of factor IX secreted from CHO cells escapes a proper cleavage of propeptide, resulting in inactive factor IX molecules with the propeptide still attached.[26] These data indicate that cultured cells such as BHK cells and CHO cells have mechanisms required for various co- and post-translational modifications with rather low, limited capacities.

Studies on the vitamin K dependent proteins have provided evidence of a specific biological role of propeptide in protein biosynthesis. The propeptide of von Willebrand factor has been shown to be required for multimerization of this protein, providing another function for such sequences.[27–29] Interestingly, the propeptide (except the first Thr residue) and Gla domain of factor IX are coded by a second exon, suggesting that these two adjacent unique domains are evolutionally one unit (Figure 2).[9] Several mutant factor IX genes containing mutations in the Gla domain,[30,31] such as factor $IX_{Chongqing}$[32] which has its Glu27 replaced with Val, provide invaluable information on the structures required for the function of the Gla domain. The Gla domain binds calcium ions with a moderately low binding affinity (average $K_d = 0.8\ \mu M$).[33,34] Binding of calcium ions to the Gla domain is required for its conformational rearrangement from a disordered form to an ordered and organized form involving the epidermal growth factor (EGF)-like domain. This conformational rearrangement is essential for factor IX to bind to negatively charged phospholipid vesicles provided *in vivo* by activated platelets resulting in its localization and augmentation of activation. Recently, the X-ray crystallographic structure of the Gla domain of prothrombin fragment 1 was determined.[35,36] This structure shows that in the absence of calcium ions, most of the Gla domain (aa 1–35) is substantially disordered. However, when the fragment 1 was crystallized in the presence of Ca^{2+}, the structure of the Gla domain was found to be well ordered, giving enough intensity of diffracted X-ray for a detailed analysis. This agrees well with the above observations obtained from experiments in solution. The Gla domain is composed of four separate short α-helices. The Gla domain of prothrombin fragment 1 binds seven Ca^{2+} ions containing four trapped between two parallel structures formed of two segments including residues 7 and 8, and residues 20, 21, 27 and 30. All Gla residues found in prothrombin are also conserved in factor IX, suggesting the similar Ca^{2+} binding may be expected for factor IX. A mutation (Gla27 replaced with Val) found in factor $IX_{Chongqing}$,[32] therefore, apparently disturbs an important Ca^{2+} binding site in the Gla domain. Furthermore, mutagenesis analyses suggest that both Gla20 and Gla21 are required for maintenance of the structure recognized by factor XIa in activation, and that Gla21, but not Gla20, is also necessary for the calcium-dependent conformational change and endothelial binding of factor IX.[37] Factor IX, however, contains two more Gla residues (aa 36 and 40) which are not shared in prothrombin. Whether or not these are also involved in extra Ca^{2+} ion binding is not known. Binding of human factor IX to endothelial cells requires a small region of the Gla domain spanning residues 3–11.[38]

Two EGF domains in factor IX do not have any growth factor-like activity[39] and may have conformations of antiparallel pleated sheets as shown for factor X.[40] Only the first epidermal growth factor-like domain contains a high-affinity calcium binding site.[33,41] Binding calcium ion to this domain is essential to initiate conformational rearrangements involving the Gla domain.[33,34] The first EGF domain

(NH_2-terminal domain, corresponding to aa 47 through 84) undergoes at least three types of post-translational modifications. These include erythro-β-hydroxylation of Asp64,[42] *O*-glycosidically linked di- or trisaccharide (D-Xyl*p*a1-3-D-Glcb1-*O*-Ser53, or one more D-Xyl extension) in human or bovine factor IX, respectively,[43] and three disulphide bond formations. β-Hydroxylation of Asp64 forming β-hydroxyaspartate (Hya), which is catalysed by a 2-oxoglutarate-dependent dioxygenase in liver microsomes, is only partial in factor IX (~30% complete). This is markedly different from other proteins such as factor X which undergoes a complete modification at this site. Dioxygenase does not require vitamin K for its activity, and β-hydroxylation is a reaction independent of γ-carboxylation.[44] By using inhibitors that block aspartyl β-hydroxylation of recombinant human factor IX, the Hya residue in factor IX was demonstrated to be non-essential for factor IX function as well as for binding to endothelial cells.

By a series of intrinsic protein fluorescent studies with various portions of factor IX, the first EGF domain was further studied for its high-affinity calcium binding site(s) (half saturation at ~40 μM Ca^{2+}).[33,34] This site is present independent of the state of carboxylation of the Gla domain. When Asp64 is not β-hydroxylated, the EGF domain still maintains a high-affinity Ca^{2+} binding site, although with K_d = 200–300 μM. Calcium ion binding at the high affinity site in the first EGF domain appears to induce significant conformational changes in factor IX that are detected by changes in intrinsic protein fluorescence, higher resistance to Lys endopeptidase and less accessibility to disulphide bonds by reducing agents. Binding of Ca^{2+} to the Gla domain, which has about ten-fold lower affinity for Ca^{2+} compared with the high affinity site in the first EGF domain, is required to complete the conformational rearrangement involving Gla and EGF domains. This rearrangement is necessary for factor IX to bind to the membrane surface. As shown for protein C,[45] the EGF domain may also affect the conformation and activity of the catalytic subunit of factor IX. In addition to the part of Gla domain sequence (residues 3–11 and 21),[38] the first EGF domain may be involved in the binding of factor IX to endothelial cells with a high affinity (K_d = ~2 μM).[33,34] The importance of the first EGF domain for factor IX function is supported by the detrimental effects of many mutations found in this domain,[30,31] such as factor $IX_{Alabama}$ containing Asp47 replaced with Gly (10% of the normal factor IX activity),[46] factor $IX_{New\ London}$ containing Gln50 replaced with Pro (< 1% activity),[47] and factor $IX_{Hollywood}$ containing Pro55 replaced with Ala (11% activity).[48] Interestingly, once activated, factor $IX_{New\ London}$ shows about 17% of normal activity which is comparable with other abnormal factor IX with mutations in the first EGF domain. The delayed activation of factor $IX_{New\ London}$ is in part responsible for its lowered activity. The replacement of Gln50 with Pro may also disrupt factor VIII binding, as observed for factor $IX_{Alabama}$ which shows a reduced effect of factor VIII on activation of factor X by factor IX.[46] Replacement of Pro55 with Ala in factor $IX_{Hollywood}$ was speculated to disrupt a β-turn structure required for the putative antiparallel β-sheet conformation of this domain.[48]

By swapping domains, Lin *et al.*[49] found that the first EGF domain of factor X can functionally replace that of factor IX, but the second (COOH-terminal side) EGF domain of factor X cannot assume the function of the counterpart of factor IX. These data together with those mentioned above suggest that the first EGF domain may function as a scaffold for holding the Gla region in Ca^{2+}-induced conformational rearrangements.[34] The second EGF domain of factor IX was also suggested to be involved in interaction with factor VIII.[49] This is further supported by a mutant protein which has a replacement of Asn92 with His in the second EGF domain.[50] Hertzberg *et al.*[51] reported that the second EGF domain and the protease domain of factor Xa in a chimera with the factor IX amino-terminal half including signal peptide, propeptide, Gla domain and the first EGF domain are sufficient to interact with factor Va. More recently, the first EGF domain was found to be required for factor IX activation by factor VIIa–tissue factor pathway, but not by the factor IXa pathway. It is also essential for optimal activation of factor X by factor IXa/factor VIIIa/phospholipid complex, but for neither phospholipid nor factor VIIIa binding to factor IXa (P. Bajaj, personal communication).

Most information on the structure–function relationship of the rest of the factor IX molecule, including the protease domain (catalytic subunit), has come from analysis of a large number of natural mutations, particularly missense mutations, which are distributed throughout the molecule.[30] The COOH-terminal side of the second EGF domain is connected to a linking region (aa 129–145). This short sequence is not clear except for Cys132 which is involved in a disulphide bond with Cys239 of the heavy chain, and Arg145 which is involved in one of the two proteolytic cleavage sites for activation of factor IX.[52] Mutant factor IX molecules with Cys132 replaced with Arg (factor IX_{Dakar})[30] and Arg145 replaced with Cys (factor $IX_{Cardiff}$ and others)[53] or His (factor $IX_{Chapel\ Hill}$ and others)[52] support the importance of these residues. During proteolytic activation, cleavages of two peptide

bonds, one between Arg145 and Ala146 and the other between Arg180 and Val181, release an activation peptide of 35 amino acid residues.[54] Structural requirements for the activation peptide appear not to be stringent except the immediate neighbouring sequences of the proteolytic cleavage sites which are involved in specific interactions with the activating enzyme. Absence of missense mutations, except the above-mentioned, in the regions of the linking and activation peptide, agrees with the notion that these regions function as spacers and are generally permissive for various amino acid sequence changes.

Mutations found in the catalytic subunit are also distributed throughout the domain. Mutations in some regions of the catalytic subunit, such as Pro287 replaced with Leu, Ala291 replaced with Pro, and Thr296 replaced with Met, apparently destabilize the protein or are detrimental for protein secretion, resulting in low-antigen, low-activity type variants.[30,31] Mutations in the highly conserved areas such as Gly363 to Val, Pro368 to Thr or Gly367 to Arg in the immediate neighbourhood of the active site residue (Ser365) cause a severe disorder, suggesting that these sequences are essential for keeping the active site structure functional. For example, factor $IX_{Eagle\ Rock}$, which has Val instead of Gly363, is apparently as stable as the normal factor IX, but cannot form a correct active site conformation because of the side chain of Val.[55]

Haemophilia Bm phenotype is characterized by its prolonged partial thromboplastin time and prothrombin time with ox (but not human) brain tissue factor,[56] and has mutations in the two distinct regions, an activation site area (Arg180–182) and another area (residues 390–397).[30,31] These mutant factor IX proteins likely interfere with the proper binding of factor VII (or VIIa) to the bovine tissue factor.

A large number of amino acid residues (about 60% of the factor IX sequence) apparently serve just as spacer sequences to maintain the overall factor IX protein structure, and are replaceable with most other amino acid residues with different side chains without resulting in haemophilia B.[57]

More recently, a high affinity Ca^{2+} binding site ($K_d = \sim 500\ \mu M$) in the catalytic subunit was reported to be possibly involved in binding of factor IXa to factor VIII.[58] Carbohydrate chains attached to factor IX may play an important role in activation or function of factor IX as observed for factor X.[59]

Structure of factor IX gene

Complementary DNA and gene of human factor IX were cloned and their complete nucleotide sequences have been determined.[8,9,60] The nucleotide numbering originally employed for the complete contiguous sequence[9] has become the standard system for the gene and is used in this article. Factor IX is composed of eight exons in a span of about 34 kilobase (kb) pairs (Figure 3).[9] The size of the gene, however, may be as large as 40 kb depending on unidentified regulatory elements located in the 5′ and/or 3′ flanking sequences. The factor IX gene is located on the X-chromosome at q27 in an order of centromere—HPRT at q26—FIX at q27.1—fragile site at q27.3—Factor VIII site at q28—telomere.[61–63]

Bottema *et al.*[64] reported that the G + C content of the factor IX gene (40%) which is in general agreement with that of mammalian genomes cannot be explained by C to T or G to A transition alone at the CpG sites. The mutation rate at CpG sites is elevated about 24- and 7.7-fold relative to other transitions and transversions, respectively. Given the enhanced mutation rates at CpG, two-fold and three-fold mutational enhancement for transitions and transversions, respectively, at

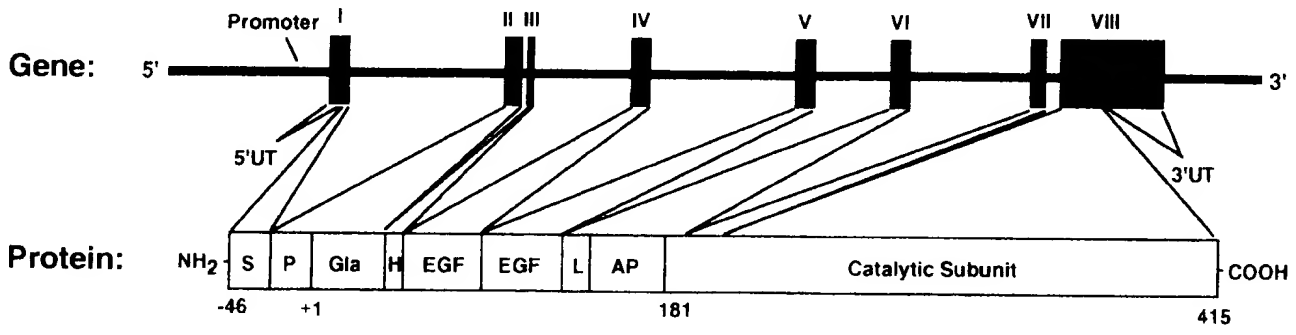


Figure 3. Organization of the human factor IX gene and domain structures of factor IX molecule. Exons are shown as solid vertical bars with exon numbers. S, P, Gla, H, EGF, L and AP represent signal peptide, propeptide, Gla domain, hydrophobic sequence, EGF-like domain, link sequence and activation peptide, respectively. Corresponding exons and domains are shown by lines. Numbers below domain structures indicate corresponding amino acid (aa) residue numbers.

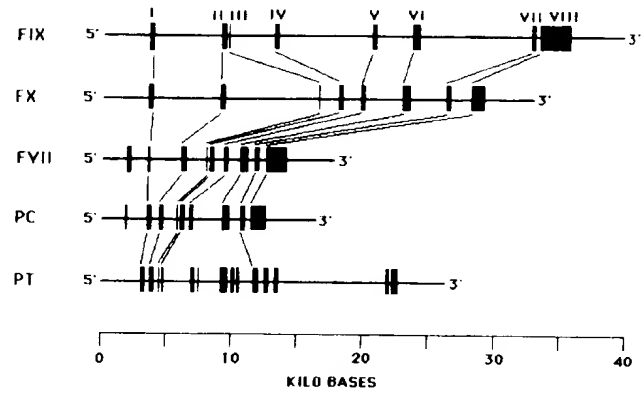


Figure 4. Comparison of gene organizations for factor IX (FIX), factor X (FX), factor VII (FVII), protein C (PC), and prothrombin (PT). Exons are depicted by solid vertical bars. Corresponding exons are shown by thin lines. Modified from Kurachi and Chen.[83]

non-CpG sites would be sufficient to produce the low G + C content of the factor IX gene. The G + C content in some other genes such as the protein C gene (over 50%) is significantly higher than that of the factor IX gene. Although the precise reason for this difference is not known, speculated reasonings may include that the specific location of each gene in chromatin which may vary in their susceptibility to base changes and/or the evolutional pressure to minimize the number of mutations in some crucial genes maintaining any functional changes at minimum.

The exon–intron organization including splicing phases of the factor IX gene are surprisingly similar to several other vitamin K dependent coagulation factors, indicating a typical divergent evolution involving a large number of point and minor mutations after exon shuffling events.[9] The overall sizes vary largely among these homologous genes due to different sizes of corresponding introns (Figure 4). Sizes of these genes range from 11 kb for the protein C gene[65] to 34 kb for the factor IX gene.[9] Protein S[66] and prothrombin[67] have only the first three exons homologous to factor IX. These exons encode its prepro leader sequence, Gla domain and a short hydrophobic sequence, but the rest of the molecules have grossly different structures. Prothrombin has two kringles and a protease domain with different exon–intron organization, while protein S has four EGF domains followed by a domain similar to steroid hormone binding protein. Protein Z has a similar organization to factor IX and other closely related factors, although it lacks two crucial amino acid residues required for the protease active site formation and does not have protease activity.[68,69] The number of introns and splicing phases in the catalytic domain of factor IX is identical to those of the genes for factor X,[70] factor VII[71] and protein C,[65] but distinctly different from those for prothrombin which has more introns in the domain.

Eight exons of the factor IX gene encode distinct domain structures (Figures 2 and 3). The 5′ end untranslated (UT) sequence and the signal peptide are encoded by the first exon. The entire propeptide (except the first Thr residue at −18 which is coded by the first exon) and Gla domain are encoded as one genetic unit by the second exon corresponding to their coordinated functions in the vitamin K dependent γ-carboxylation of glutamic acid residues in the Gla domain.[9] Other introns are also present at positions separating various unique domains.

Currently, more than ten polymorphic sites have been identified in or near the human factor IX gene (Table 1). Most of these are identified as changes in restriction sites including Bam HI site in the 5′ immediate flanking region (nucleotide sequence: C or T),[72] Hinf I/ Dde I site in a 50 bp AT-rich insert in the first intron,[73] Xmn I in the third intron (G or C),[73] Taq I site in the fourth intron (C or T),[74] and Mnl I site in the sixth intron (originally identified as Thr/Ala dimorphism at aa 148).[75] A highly polymorphic site (A > G) in Japanese population was found in the first intron.[76] An intragenic Bam HI polymorphic site located at −500 bp 5′ to the Xmn I polymorphic site in intron 3 in about 50% of the African American population.[77] Msp I polymorphism in strong disequilibrium with the Taq I polymorphism in the fourth intron was also reported.[78] Most intragenic polymorphic sites are in strong linkage disequilibrium, and these polymorphic alleles co-segregate in 70–80% of Caucasian factor IX genes. The Bam HI polymorphic site in the 5′ end region and Hha I polymorphism found at 8 kb downstream in the 3′ flanking sequence[79,80] are in equilibrium. Extragenic polymorphic Sst I site (detected with DXS99) and Taq I site (detected with DXS102) in the 5′ upstream flanking sequence and the factor IX gene may have a 3–5% chance of recombination due to cross-over events in meioses.[81,82] Linkage disequilibrium among the intragenic polymorphic sites and the possibility of recombination for the extragenic sites significantly hamper the usage of these for carrier detection and prenatal diagnosis. The overall usefulness of these polymorphisms in carrier detection is about 90% for whites and blacks. Frequencies of these polymorphisms vary largely among ethnic groups.[83] Except for the A/G polymorphic site in the first intron,[76] all intragenic restriction fragment-length polymorphisms which are present in Caucasians and African Americans are absent or extremely rare in Asians.[83]

Besides the polymorphic AT-rich sequence in the first intron, tandem purine and pyrimidine dinucleo-

Table 1. Allele frequencies of polymorphisms of human factor IX gene in various populations[a]

Restriction enzymes	Position	Allele (kb)	Caucasian		African American		Asian	
			No. of chromosomes	Frequency	No. of chromosomes	Frequency	No. of chromosomes	Frequency
Sst I	5′ extragenic	6	76	0.48	38	0.53	63	0.5
		9	81	0.52	34	0.47	63	0.5
Taq I	5′ extragenic	10	18	0.90				
		1.2	2	0.10				
Bam HI	5′ extragenic	25	90	0.94	50	0.64	100	1.0
		23	6	0.06	28	0.36	0	0.0
Hinf I/Ddel	intron 1	1.75	10	0.21	14	0.36	0	0.0
		1.70	36	0.79	26	0.64	30	1.0
Bam HI[b]	intron 3	15	32	1.0	11	0.52		
		13	0	0.0	10	0.48		
Xmn I	intron 3	11.5	42	0.7	12	1.0	76	1.0
		6.5	16	0.3	0	0.0	0	0.0
Taq I	intron 4	1.8	285	0.68	15	0.9	61	1.0
		1.4	129	0.32	2	0.1	0	0.0
Msp I	intron 4	2.4	40	0.8	22	0.4	57	1.0
		5.8	10	0.2	31	0.6	0	0.0
Mnl I[c]	exon 6	(Ala)	10	0.29	8	0.12	0	0.0
		(Thr)	25	0.71	60	0.88	95	1.0
Hha I	3′ extragenic	0.2	13	0.38	11	0.33	43	0.83
		0.15	21	0.62	22	0.67	9	0.17
purine/		I	4	0.29			0	0.00
pyrimidine		II	10	0.71			13	0.93
polymorphism[d]		III	0	0.00			1	0.07

[a] The data summarized in this table is a composite of data reported in papers including Hay *et al.*,[72] Winship *et al.*,[73] Camerino *et al.*,[74,78] Driscoll *et al.*,[77] Reiner *et al.*,[80] Mulligan *et al.*,[81] Hofker *et al.*,[82] Kurachi *et al.*,[83] Sarkar *et al.*[84] and personal communications (S.-H. Chen).

[b] This 2nd *Bam* HI site is located at 500 bp 5′ to the *Xmn* I polymorphic site and was detected by a *Bam* HI/*Sph* I digest.

[c] This is known as Thr/Ala-148 dimorphism which codes either for Thr or Ala.

[d] Allele I, $(GT)_6ATGC(GT)_4AG(AC)_4GCAT(AC)_3$; Allele II, $(GT)_5TGC(GT)_4AG(AC)_4GCAT(AC)_3$; Allele III, $(GT)_5ATGC(GT)_4AG(AC)_3GCAT(AC)_3$.

tide repeats which are polymorphic in most human races are also present in the 3′ UT sequence in the form of four different alleles.[84] These polymorphisms further improve carrier determination and prenatal diagnosis of haemophilia B.

The human factor IX gene contains many repetitive sequences such as Alu sequence and long interspersed element (Kpn I repetitive sequence, abbreviated as Line-1 or L-1). Five Alu sequences are present in introns and in the 3′ end immediate flanking region, while two Line-1 sequences, one partial element and one 6.1 kb complete element, are present in the fourth intron and in the 5′ flanking region, respectively.[9] Novel, short, interspersed repeat sequence (Ano) is also present in the first intron.[85]

Abnormal factor IX genes

To date, more than 600 haemophilia B patients have been studied for molecular defects in their factor IX genes.[30] Of 574 patient entries in the 1992 database (Third edition), 278 (48%) are unique and the rest are repeats which may be due to independent mutations or founder effects.

Mechanisms of mutation found in haemophilia B genes are highly heterogeneous, including at least 29 cases of complete or partial gene deletions or more complicated gene rearrangements, 50 short (less than 20 nucleotide) deletions or insertions or both, and a large number of single-base mutations which include 524 cases of missense, nonsense mutations, and mutations at splice sites as well as in the 5′ UTR and flanking sequence.[30,31] Gross and relatively large gene deletions, insertions and rearrangements, which can be rapidly detected by Southern blot analysis or polymerase chain reaction (PCR) as missing or rearranged DNA fragments, account for only about 4% of all mutations. Some of the gross gene deletions may be parts of much larger deletions which may span more than 500 kb in size.[31] All patients with gross gene deletions are severely affected. However, only two-thirds of those patients have developed alloantibodies

(inhibitors) against human factor IX infused during replacement therapy. Some patients with detectable factor IX antigens also develop alloantibody. These observations indicate that the development of inhibitors against the normal factor IX infused in the protein replacement therapy is primarily due to secondary factors such as treatment regimens and/or polymorphisms in the immune response system of individuals, but is not simply due to missing the entire factor IX antigen or to specific epitopes resulting from gross or partial gene deletions.

No obvious hot spots for deletion breaking points have been observed. However, factor $IX_{Seattle\ 1}$ which has about a 10 kb deletion spanning introns 4 through 6 has been shown to involve the 14 bp sequence (TAGAAGTTCACTT) duplicated 10 kb apart in introns 4 and 6.[86] In some cases, well-known repetitive sequences such as Alu sequences (highly abundant repetitive sequence of about 300 bp in length) are apparently involved.[87] An interesting mutant gene with an insertion is factor $IX_{El\ Salvador}$.[88] This has an insertion of about 6 kb extra sequence (which is likely an L-1 element) in the fourth intron within the 0.8 kb which spans between the 3′ end of exon 4 and the first Eco R1 site in this intron. Whether or not the insert sequence is directly responsible for this haemophilia B is yet to be determined. Line-1 element has a complete transcriptional unit with two open reading frames including retroviral reverse transcriptase-like sequence.[88] Line-1 insertion in exon 14 of the factor VIII gene has been reported as a novel mutational mechanism.[89] Line-1 sequence inserted in an intron may possibly generate an extra splicing set of sequences causing abnormal processing of the factor VIII gene in a mild haemophiliac.[90] Otherwise, Line-1 element insertion in introns may not be deleterious, and haemophilia in this type of kindred is incidental, probably due to an unidentified second mutation somewhere else in the gene.

About 40 mutant factor IX genes with small deletions such as one, two, three, four, seven and more than 13 bases in size have been found distributed in exons as well as in introns.[30,31] No specific hot spots for these small deletion mutations have been observed. Point mutations found in abnormal factor IX genes are distributed throughout the factor IX molecule, suggesting that the entire structure of the factor IX protein is highly optimized and that almost every part of the molecule is essential for maintaining its overall structure and/or specific function. A large part of the sequence is estimated to serve as spacers for maintaining the overall globular structure, as mentioned above.[57] This may be different from factor VIII with its dispensable, large central B domain. Currently, no function, other than its function as a large spacer (activation peptide), has been identified for the B domain of factor VIII.[91,92]

Missense mutations account for the majority (70%) of the point mutations in abnormal factor IX genes, while nonsense mutations account for about 16%.[30] These mutations, particularly missense mutations in factor IX genes, result in subtle changes of factor IX structure causing a wide spectrum of clinical severity and providing us with insights into the structure–function relationship of factor IX. Because of its relatively small size, availability of a large number of mutant genes with missense mutations and the complete gene structure, factor IX may be the most amenable protein among all coagulation factors for the exhaustive study of structure–function relationships.

Mechanisms responsible for the point mutations found in the factor IX gene are highly heterogeneous. Among them, CpG dinucleotide sequence has been clearly recognized as a mutational hot spot.[30,31] Eighteen (or 36 for the double strands of DNA) CpG sequences are present in the coding region, which is only a quarter of the possible random dinucleotide sequences. Endogenous methylase converts some of deoxycytidine of the CpG sequences to 5-methyldeoxycytidine which is then spontaneously converted to deoxythymidine by deamination. Because no cellular repair mechanisms are present for this conversion, the rate of mutation at CpG in factor IX genes is elevated at least 24-fold and 7.7-fold for transition and transversion, respectively, over the other random mutations at non-CpG sequences.[64,93,94] Interestingly, the lack of repair mechanism for this alone cannot account for the reduced frequency of CpG sequences in the gene.[64] About 45% of all unique point mutations found in factor IX genes are due to mutations at CpG sites. Some of these mutations observed may be due to founder effects. Data obtained on factor IX genes agree well with those for other genes such as the factor VIII gene.[95] Twelve CpG sites in the factor IX gene have base mutations once or multiple times. Mutations at CpG sites such as Thr296 replaced with Met are duplicated in unrelated families, further supporting these CpG sites as hot spots.[96] Some CpG sites do not have any mutations reported. The underlying mechanism for this is not known. It is possible, however, that these CpG sites are somehow inaccessible to methylase in the chromatin structure. Multiple different mutations at the same nucleotide sequences of CpG or at non-CpG sites have also occurred.[30,31] Some examples include replacements of nt −6G (5′ UTR) by A or C, nt +13A (5′ UTR) by G or C, or deletion, nt 6 365G in a codon for Arg-4 by T or A, nt 6 704T (splice site) by G or C, nt 10419G in a codon for Cys56 by C or A,

20524G in a codon for Val182 by C or T, nt 20566G (splice site) by T or A, nt 30992G in a codon for Ala291 by C or A and nt 31290C in a codon for Ala390 by T or A. These data further support the concept that the mechanisms responsible for the mutations are highly heterogeneous, and demonstrate that the factor IX structure has been extensively tested and refined by mutational events in the process of evolution.

An interesting case of somatic mosaicism of abnormal factor IX with a mutation (Cys350 to Ser due to G to C change) has been reported.[97] In a family affected with haemophilia B, a male member was very mildly affected (35% factor IX activity with 45% antigen level of normal). In this haemophilia B kindred, the two female members (daughter and granddaughter of the affected male) were moderately affected (3% activity and 4% antigen of the normal factor IX level). Factor X and factor VII levels as well as prothrombin time were normal. Among somatic tissues of the affected male analysed, about 10% of the total cells of both leucocytes and liver have the normal gene while about 90% of the cells have the mutant gene. However, the cells of kidney and smooth muscle have both normal and mutant genes about equally. These results indicate a somatic mosaicism, probably due to a replication or post-replication repair error during the first mitotic divisions in the zygote preceding implantation, or a half-chromatid mutation generated during meiosis that was not corrected before fertilization. In the leucocytes of the two female patients, both normal and mutant genes are present in an equal amount. These data suggest a possibility that not only liver but also leucocytes are of endoderm origin, which is contrary to the commonly held mesodermal origin for leucocytes. A rare case of severe haemophilia B in a girl due to non-random inactivation of a normal factor IX gene has also been reported.[98]

A class of abnormal genes which have mutations in the Leyden-specific region (LS-region, arbitrarily defined as a region from about nt −40 to +20) belong to the haemophilia B-Leyden phenotype, which shows a unique delay of the factor IX expression until the onset of puberty. Eighteen mutations have been found in the 5′ end region of the factor IX gene. Among these, one is located at nt −793 in the 5′ upstream, and the rest (17 mutations found) account for twelve unique mutations in the LS-region. Eleven of these mutations result in the Leyden phenotype.

Regulation of the factor IX gene expression

The factor IX gene is expressed in liver with a high tissue specificity.[99] Illegitimate expression in other tissues is also observed.[100]

During most of the gestational period, the factor IX

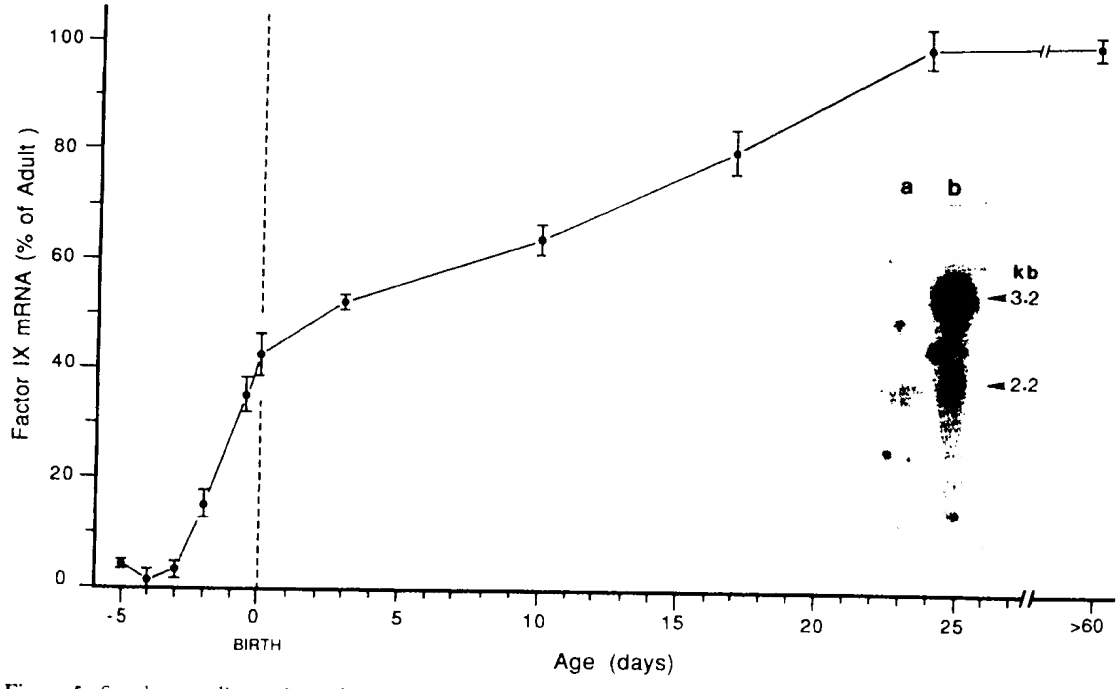


Figure 5. Steady-state liver mRNA levels for mouse factor IX at various developmental stages. Factor IX mRNA levels (solid dots) are shown as percentages of that of the adult. Vertical bars with short horizontal bars indicate standard errors (*n*: 6–8). The inset picture shows the results of Northern blot analysis of poly(A)⁺RNA prepared from NIH/3T3 cells (lane 1) and mouse liver (lane 2). The numbers on the right indicate the sizes of the estimated transcripts. Taken from Yao *et al.*[103] with permission.

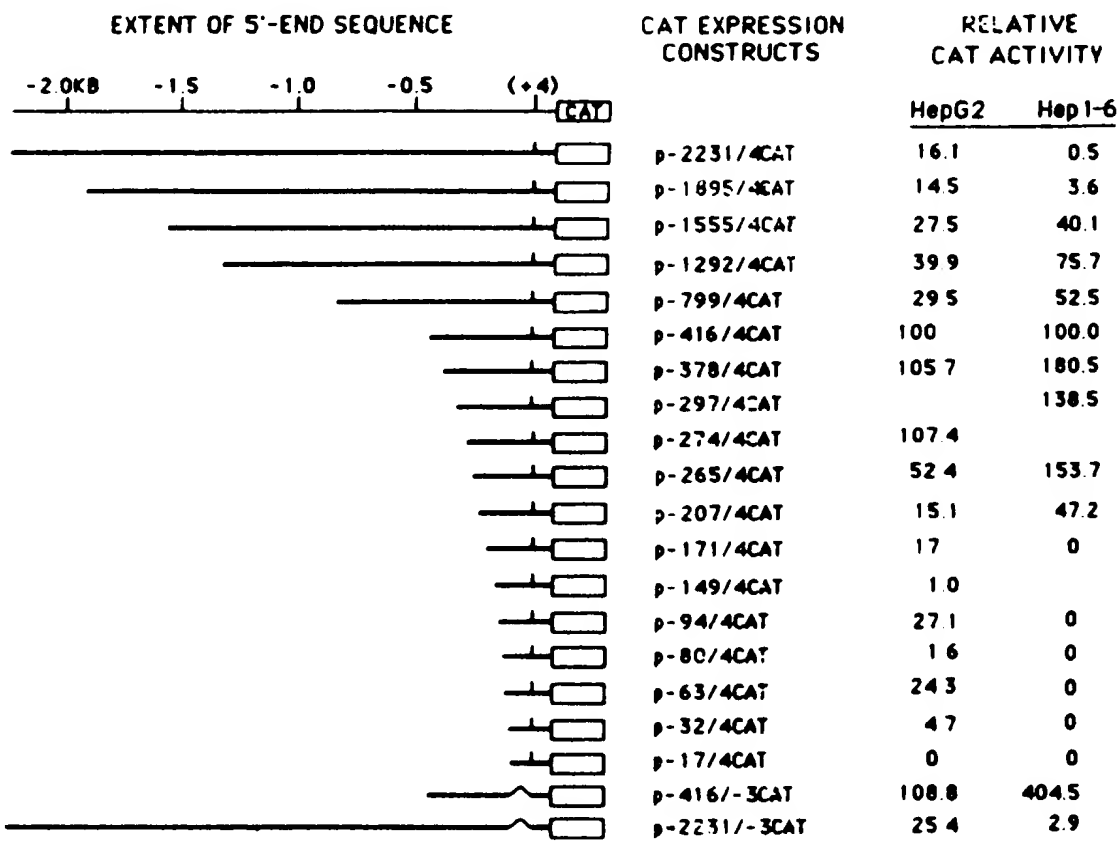


Figure 6. Analysis of the 5′ end flanking sequence of the factor IX gene for promoter activity. Expression activities of CAT constructs containing various portions of the 5′ end flanking sequence were assayed in HepG2 cells and Hep1-6 cells.[99] The 5′ end sequence of the factor IX gene is shown with solid lines with the CAT gene shown as an open box at the 3′ end. CAT activities relative to that of p-416/4CAT are shown in percentages. The thin lines connected in the middle indicate deleted areas. The 5′ and 3′ ends of the factor IX gene sequence contained in each CAT construct are shown by two numbers separated by a slash in the labellings.[9] p-416/29CAT, which contains a factor IX sequence extended to nt +29 at the 3′ end, shows identical expression activity to that of p-416/4CAT. Deletions of the 3′ end sequence from nt +29 up to −2 in p-416/29CAT did not effect expression activities of CAT constructs. However, further deletions beyond nt −3 position dramatically reduced the activity.

gene is expressed only at a low level (3–5% of the adult level) until the late stage of the third trimester. This was shown for humans, with limited data,[101] or lamb[102] and more completely for mice.[103] The developmental time curve of expression of the factor IX gene in mouse liver shows an induction of a high-level expression of the factor IX gene on day 18 of gestation (late stage of the third trimester) (Figure 5). The increased expression of the gene continues through birth followed by a rather gradual increase until reaching the adult level at weaning (20–24 days of age). At birth, the factor IX mRNA level is only 43% of the adult level, and the plasma factor IX activity level agrees well with the mRNA level. These results agree well with the limited data available on humans.[104,105] The low level of factor IX mRNA at birth may be responsible in part for haemorrhagic disorders in pre-term or term neonates. This condition, however, may be aggravated by generally poor vitamin K synthesis in neonates and, furthermore, if antibiotics or vitamin K analogues such as warfarin are given to the mother during the prenatal stage.[106] Other pathological conditions such as diarrhoea and cystic fibrosis also lower the vitamin K level, resulting in secondary haemorrhagic diseases.[107]

As in any other gene, the 5′ end region of the factor IX gene contains various short sequences which function as *cis*-acting elements in its regulation.[99] Systematic analyses of these elements in the 5′ end region have been carried out with expression vector constructs containing variously deleted 5′ end region of the factor IX gene ligated to chloramphenicol acetyltransferase (CAT) gene as a reporter (Figure 6).[99] The schematic drawing of the overall organization of the major functional elements identified is shown in Figure 7. The

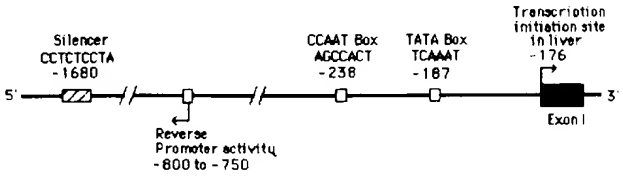


Figure 7. Locations of major functional elements in the 5′ end region of the factor IX gene. Nucleotide numbering system is from Yoshitake *et al.*[9] The nucleotide sequence of the silencer is of the complementary strand.

fundamental elements necessary for a high-level expression of the factor IX gene are contained in approximately the first 300 bp sequences of the 5′ end region (Figure 7). As more 5′ upstream sequences beyond nt −400 region are included into the CAT constructs, lower expression activities are observed (Figure 6). The sequence including up to about nt −1900 shows only a low-level activity (16% of the construct with a sequence up to nt −416). Even lower expression (less than 3%) is observed when the 5′ upstream sequence up to −6.9 kb is included in the expression vector. The factor IX gene does not have a typical TATA sequence in the 5′ end flanking sequence. However, according to the functional analysis data, sequences TCAAAT starting at nt −187 and AGCCACT starting at nt −238 have been tentatively identified as functional TATA box and CCAAT box, respectively. AGCCACT agrees well with the consensus CCAAT sequence.[108]

The locations of fundamental transcriptional elements of the factor IX gene agree well with the transcriptional start site placed in the region of nt −150 for CAT constructs[99] which was revised from the previously assigned site (+1 site).[109,110] The primary transcription start site in liver was determined to be at nt −176 in the 5′ upstream by primer extension and DNase I protection analyses with high-quality poly (A)⁺RNA preparations of human livers.[111,112] Reverse transcription-mediated PCR with primers at or downstream of nt −176, but not with one at nt −300 region where no signals for transcription initiation were observed, can amplify products, further supporting the

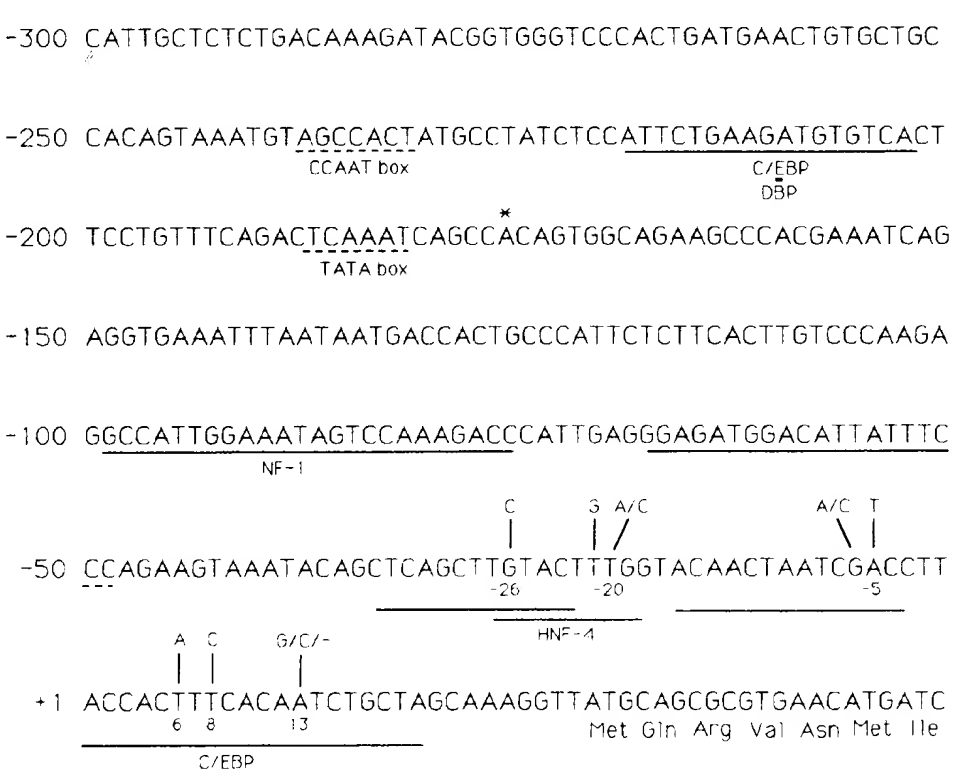


Figure 8. The nucleotide sequence of the 5′ end region (300 bp) with various elements. The numbers on the left indicate the old nucleotide numbering.[9] The revised primary transcription start site in liver is shown with an asterisk. Part of the signal peptide starting with the first Met residue at aa −46 is shown at the bottom line. Solid underlines indicate protein binding regions. Dotted underlines indicate tentatively identified functional CCAAT and TATA boxes. The LS-region is arbitrarily defined as the region roughly spanning nt −40 to +20. Mutations found in haemophilia B_{Leyden} genes at nt −20, −6, −5, +6, +8 and +13 are shown with short vertical bars with mutant sequences above the LS-region. The mutation at nt −26 (G to C change) does not show a Leyden phenotype.

initiation site in liver. A dog factor IX cDNA which has the 5′ extension at least up to −179 is also in good agreement with the revised 5′ upstream start site for the human gene.[113] The previously observed transcription initiation site (+1) is likely a secondary start site in liver or an artifact, probably due to poor quality of the RNA preparations used. The site, which is located in the middle of a region designated as LS-region, could be in a unique secondary structure which makes factor IX mRNA highly susceptible to degradation at this site. Crossley and Brownlee[114,115] reported functional analyses of the 5′ end promoter region for transcriptional activity using a CAT construct containing a 5′ end sequence up to nt −189 as a control for the full CAT activity. This construct, which contains the factor IX sequence up to nt −189, however, has only about 15% of the optimal constructs which contain a 5′ sequence up to nt −300.[99]

Structural elements homologous to the known liver-specific enhancers are also present in the 5′ end promoter region. These include TGGACC (partial LF-A1 or HNF-4 element) at nt −359 in sense strand and CTTTGGACT (PRI element) at nt −79 in antisense strand, which are also present in other genes such as α_1-antitrypsin, transferrin and antithrombin III genes.[99] The region, nt −76 through −99, containing CTTTGGACT has been reported to bind NF-1[114] which is originally identified as a liver-specific enhancer protein.[116,117]

Several negative regulatory elements (silencers) which are identical or similar to those found in other genes are present in the region of about nt −700 through −2000. These elements are responsible for the activity reduction observed for CAT constructs which contain various portions of sequence in this region.[99] Among them, a sequence spanning −1.4 kb to −1.7 kb contains two sequence elements, ATCCTCTCC starting at nt −1680 and CAATGGTT at nt −1621, which are similar to the well-characterized consensus silencer elements (negative regulatory elements).[99,118] When a sequence containing these elements was placed downstream of a CAT gene at the Bam HI site in p-416/+4CAT which contains the factor IX promoter sequence (nt +4 through −416), both orientations of this sequence (sense and reverse orientation) reduced the expression activity to 21–26% of p-416/+4CAT.[99] These results indicate that silencer elements in the region are actually functional in the factor IX gene. More silencer-like elements found in the 5′ end region include ACCTATGGAA starting at nt −726, CTGAATGGCT at nt −793 and CAATGACT at nt −1467. Interestingly, a very strong promoter activity in a reverse direction is present in a region spanning nt −700 to −750.[99] The sequence elements responsible for this reverse direction promoter are currently not known. No retroviral LTR-like sequences are present in this region. The presence of the reverse promoter region appears to coincide with a significant reduction (60–70%) of the normal expression activity of the factor IX gene (Figure 6).[99]

Important information on the regulation of the factor IX gene has also been obtained from transgenic mice experiments. Jallat *et al.*[119] have recently constructed transgenic mice carrying factor IX minigenes with the 5 kb sequence of the 5′ end immediate flanking region containing the promoter elements in addition to all the silencer elements detected in the *in vitro* assay and variously shortened intron sequences. Their results have clearly indicated that the liver-specific high expression of the factor IX gene can be achieved by various constructs with the 5 kb 5′ end flanking sequence as the promoter and the partial sequence of the first intron. A factor IX cDNA construct (containing no intron sequences) with the same 5 kb sequence of the 5′ end flanking region shows only a background level expression in transgenic mice. These observations strongly suggest that at least one intron as a set of splicing sequences or a putative enhancer element(s) which may be present in the first intron must be responsible for obviating the silencer activity in the 5′ end upstream sequence.

The data obtained from both *in vitro* and *in vivo* experiments indicate several important points, including: (i) high-level expression of the factor IX gene can be achieved by the elements contained within the sequence up to about nt −300, (ii) this expression activity is efficiently suppressed by multiple silencers present in the 5′ upstream region, and (iii) the reduced activity may be restored to a high level *in vivo* in the presence of the first intron partial sequence. The obviation of the silencer activity in the 5′ upstream sequence by the first intron sequence, therefore, appears to be a key mechanism underlying the overall regulation of the factor IX gene.

Important observations regarding the developmental regulation of the factor IX gene have been obtained from a unique class of haemophilia B, haemophilia B-Leyden.[120–122] While the normal factor IX gene is induced for its high-level expression at the perinatal stage,[103] Leyden phenotype factor IX genes are not expressed or not induced for their high expression until the onset of puberty.[120] Eleven unique single-base mutations so far found in haemophilia B-Leyden families include nt −21(T to G), −20(T to A or C), −6(G to A or C), −5(A to T), +6(T to A), +8(T to C) and +13(A to G or C or deletion).[30,121–129] Without any exceptions, all these mutations are contained in the LS-region (roughly from nt −40 to +20) in the 5′ untrans-

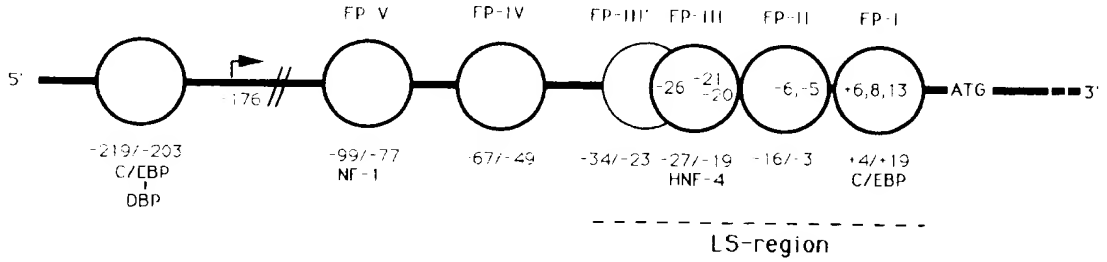


Figure 9. Protein binding at the LS-region and its neighbouring region. The human factor IX sequence is shown by a thick solid line. Proteins which bind to the region as evidenced by footprinting analyses are shown by circles. A circle with thin line indicates the protein binding in the factor IX-Leyden gene with a mutation at nt −20. Numbers in circles indicate the nucleotide residue locations of the natural mutations found in the Leyden phenotype factor IX genes. Numbers below circles indicate actual sequence regions of footprints. Known proteins are also shown below the numbers. An arrow indicates the primary transcription site in liver.

lated sequence of the factor IX gene (Figure 8). With the revised primary transcription initiation site of the factor IX gene,[99,111] the LS-region is located within the 5′ untranslated sequence. When mutations found at nt −20(T to A), −6(G to A) and +13(A to G or deletion) were tested for their effects using CAT expression vectors, all mutant sequences substantially reduced the expression activity of these constructs to 15–30% levels of the normal sequence,[112,122] indicating the importance of these sequences for the factor IX activity. As found by DNase I footprinting analyses, the normal LS-region and its neighbouring region bind five proteins[111,130] (schematically shown in Figure 9). These include three apparently new proteins, shown as footprints FP-II, FP-III′ and FP-IV,[111,130] in addition to recently reported HNF-1, C/EBP and HNF-4.[114,115,129] In the 5′ side neighbouring region, NF-1 and an unidentified new protein bind at a region spanning nt −99 to −76, and at a region spanning nt −67 to −44 (FP-IV), respectively.[130] Neither one of these new proteins is glucocorticoid receptor, nor androgen receptor.[99,111] Binding of C/EBP, which is present in most differentiated cells at significant levels, to a region spanning nt +3 to +19, is easily detected by both DNase 1 footprinting analysis and electrophoretic mobility shift analysis.[111,114,130] Recently, the protein binding to −20 region (FP-III, nt −17 to −27) was determined to be HNF-4, a protein of the steroid receptor superfamily.[115,129] HNF-4 binding to this region of the normal factor IX gene prohibits binding of another protein which binds to an overlapped androgen responsive element-like sequence (FP-III′, nt −36 to −22).[111,130] This protein, however, can competitively bind to the region of the factor IX gene if the gene has a mutation at nt −20 which causes a gross decrease in the binding affinity of HNF-4 to the region. This mutation-dependent competition between the two overlapped regions for protein binding was shown by both gel mobility shift assay and DNase I footprinting analysis.[111,130] The importance of HNF-4 binding to the region and the unidentified protein which binds to FP-III′ has been further supported by a drastic decrease in the expression activity of the factor IX gene with a mutation at nt −26.[115,129] Although Crossley *et al.*[115] reported that the protein which binds to the FP-III′ region is androgen receptor (AR), careful protein binding analyses have shown that it is not androgen receptor, but an ubiquitous protein present in nuclear extracts of liver as well as cultured cells which are not only androgen receptor-positive, such as T47D, LNCaP and HepG2, but also androgen receptor-negative cells such as CV1 and COS cells.[111,130] Glucocorticoid receptor does not bind to this region in good agreement with the results from the expression assay. When the Leyden phenotype mutations at nt +13, −6 or −20 are present in oligonucleotide sequences (double-stranded form) used in the electrophoretic mobility shift assay, the binding affinity of these proteins to the oligonucleotides is grossly decreased, agreeing well with reduced expression activities observed for the mutant sequences.[111,130] Interestingly, the 3′ half of the LS-region, where C/EBP binds at +13 subregion (FP-I), and the 5′ half, where two proteins (HNF-4 and an unidentified protein) bind in the region containing −20 and −6 subregions (FP-III and FP-II), apparently function with little cooperation.[122] The 3′ half, for instance, where C/EBP binds, may require a second unidentified element in the 5′ upstream which is not included in the CAT construct used in the assay. In this regard, Picketts *et al.*[131] recently reported that DBP interaction with C/EBP which binds at nt −202 to −219 region may synergistically confer its enhancer activity on a factor IX-Leyden gene and is responsible for the amelioration of haemophilia B-Leyden with a mutation at nt −5. DBP is induced for its expression in adulthood, but

not in childhood. This is an attractive mechanism to explain, at least in part, the Leyden phenotype. However, it has difficulties in explaining some important aspects of the Leyden phenotype. These include: (1) If DBP binding at the 5′ upstream can override the defects in the LS-region, why does DBP not ameliorate the mutation at nt −26, which is in the LS-region, after puberty? (2) Why does the normal factor IX gene which has the normal LS-region not significantly elevate its expression level after puberty as DBP increases its level? Whether or not DBP can selectively interact with any proteins, including C/EBP which binds at the 3′ half of the LS-region, is not known.

The LS-region of the transcript of the factor IX gene may assume some secondary structures such as stem loop structures, albeit not extensive, as predicted from the sequence.[122] The functional significance of stem loop structures in the untranslated region of transcripts has been well documented for other genes such as Tar element in HIV-1 or iron responsive elements in ferritin and transferrin genes.[132–134] A possible involvement of these unique structures of the LS-region in its function remains to be determined.

Development of alternative therapies for haemophilia B

Currently, haemophilia B is treated by plasma protein replacement therapy.[135] This therapy is effective, but exposes patients to possible risks of serious side-effects and complications such as contracting blood-borne viruses including hepatitis and HIV-1 viruses, thrombosis due to other coagulation factor contaminations, and inhibitor (alloantibody) development.[136] A large number of haemophilia patients (70–90%) who have received repeated plasma protein replacement therapy are already infected with HIV-1 viruses. In addition, frequent transfusions of factor IX preparation required in the therapy for severely affected patients are highly costly and significantly impair the quality of life of patients.

Large-scale production of recombinant human factor IX for safer protein replacement therapy by cultured mammalian cells is currently hampered by complicated post-translational modifications, such as γ-carboxylation, required for normal factor IX function.[25] Poor efficiency of such modification by cultured cells has been a serious problem in producing in quantity recombinant factor IX with a high specific activity. Unexpectedly, coexpression of cloned γ-carboxylase and factor IX cDNA did not improve γ-carboxylation of recombinant factor IX.[23] Information obtained from these studies, however, should eventually help to prepare recombinant mammalian cells which can express fully carboxylated factor IX. Such cells may be successfully used to produce much safer recombinant factor IX in quantity to substitute the plasma factor IX preparations currently in use.

A novel approach for an alternative haemophilia therapy includes somatic gene therapy.[137,138] This approach requires an *ex vivo* or *in vivo* transfer of the normal human factor IX gene (factor IX minigene constructed with the cDNA are widely used) into a target tissue of a patient, such as liver where the factor IX gene is normally expressed, or other tissues which can support long-term production of biologically active recombinant factor IX without any deleterious effects. If such an approach is developed, it may be able to obviate several serious side-effects of the current plasma replacement therapy.

Several cell types including rodent and haemophilic dog skin fibroblasts,[139–141] endothelial cells,[142] liver hepatocytes,[143] skeletal muscle cells,[144] and keratinocytes[145] have been tested for their ability to produce biologically active recombinant factor IX in culture. Reported recombinant factor IX preparations produced in these approaches have varied in their specific activities (~70–100% of the plasma factor IX). The variations are, in part, due to artifacts of the methods used to quantitate the recombinant factor IX secreted into medium. This problem, however, was recently solved by introduction of a simple pretreating procedure using serum with barium sulphate.[146]

When genetically modified skin fibroblasts were implanted *in dermis* or subcutaneously in mice or rats, recombinant factor IX was transiently expressed. Palmer *et al.*[140] reported that by using recombinant retroviruses containing cytomegalovirus promoter or retroviral long terminal repeat promoter, the recombinant human factor IX was produced at very high levels (~3.4 or 1.6 $\mu g/10^6$ cells/day in normal human diploid fibroblasts or in normal rat diploid fibroblasts) in culture. When these genetically modified cells were implanted into nude mice or rats, transient systemic levels of recombinant factor IX reached 0.18 μg and 0.022 μg/ml plasma, respectively. St Louis and Verma[139] originally reported systemic delivery of recombinant factor IX at a transient level of ~0.1 μg/ml serum in mice by implanting genetically modified mouse skin fibroblasts embedded in collagen under epidermis. A very inefficient systemic delivery of the produced recombinant factor IX (2–6%) as well as promoter inactivation and poor stability of the promoter used, were observed. These problems obscured the advantage of using skin fibroblasts for this purpose. Scharfmann *et al.*,[147] however, reported that use of housekeeping gene promoters such as dihydrophorate reductase may overcome some of

these problems. Recently, human applications of skin fibroblast gene therapy have been reported from China.[148] The MoMLV retroviral expression vector with its LTR as the promoter was used in these applications. One of the two mildly affected patients who received the therapy has shown a limited, transient improvement over several months. This approach still needs a substantial amount of systematic testing for its efficacy and safety before being applied to humans in this country.

Fully active recombinant human factor IX can be produced by rat capillary endothelial cells in culture at a level of 0.84 μg/10^6 cells/day.[149] A brief account of factor IX production by bovine adrenocortical endothelial cells is also reported.[141] These results indicate that endothelial cells have all the basic properties necessary to serve as a drug delivery vehicle for producing recombinant factor IX.

Skeletal myoblasts as an efficient gene transfer vehicle to obtain a high-level production of recombinant factor IX in the systemic circulation (~1 μg/ml in C3H mice) have been described.[143,145] Several important findings in this series of work include: (1) skeletal muscle cells can efficiently express foreign genes including factor IX at a high level; (2) skeletal muscle cells have mechanisms for post-translational modification producing human factor IX with a very high specific activity (81–90%); (3) efficiency of the systemic delivery of recombinant factor IX by muscles is surprisingly high ($\geq$ 29%); (4) long-term expression *in vivo* can be achieved; (5) intramuscular implanted myoblasts can, not only fuse to host myofibres, but also survive as quiescent muscle precursor cells (muscle stem cells, presumably as satellite cells), further supporting the rationale to utilize myoblast-mediated gene transfer for developing a long-term stable gene therapy for haemophilia B. Extensive efforts targeting liver are in progress in multiple laboratories. Expression of human factor IX (0.071 μg/10^6 cells/day) by rabbit hepatocytes using retroviral vector containing cytomegalovirus promoter has been reported.[143] Ponder *et al.*[150] recently reported that implantation of bacterial β-galactosidase gene-tagged hepatocytes obtained from transgenic mice (C57BL/6) by intrasplenic injection resulted in deposition and long-term survival (> 6 months) of the transplanted cells in parenchyma, which amounts to 0.5% of the entire liver. A high-level systemic delivery of human α_1 antitrypsin (~5 μg/ml plasma) in a similar approach was also observed,[151] strongly suggesting that this approach may be feasible for developing somatic gene therapy for haemophilia B. The direct factor IX gene transfer into rat liver by receptor-mediated gene transfer has shown a transient expression of biologically active factor IX into circulation.[152] The short-lived expression of factor IX observed, however, must be much improved for developing a clinically acceptable gene therapy protocol for haemophilia B.

More recently, *in vivo* expression of factor IX by taking a route of *ex vivo* gene transfer with retrovirally transduced keratinocytes was reported.[145] The expression level was extremely low ($\leq$ 1–2 ng/day/10^6 cells) and lasted less than a week, suggesting a need for more improvement before this procedure can serve the purpose. Low level expressions of dog factor IX (~6 ng/ml plasma) was also observed in partially hepatectomized haemophilia B dogs after infusion of factor retroviral vector.[153]

Currently, none of the approaches under investigation is ready for clinical testing for haemophilia in the United States. Within the next 1–3 years one or more of these approaches may be highly optimized for efficacy and safety, and become feasible for clinical applications.

Conclusion

To date, over 600 abnormal factor IX genes have been studied for their molecular mechanisms. This extensive study of factor IX in recent years is largely due to its clinical importance, the availability of its complete, contiguous nucleotide sequence which was determined in 1985, and development of readily usable technologies such as polymerase chain reactions. Furthermore, its multidomain structure with an amenable size for various protein chemical and recombinant DNA manipulations has made factor IX an exciting model for studying structure–function relationships of complex proteins.

With the enormous amount of data accumulated, new important directions of research on factor IX in the future appear to be regarding its *in vivo* role in the regulation of thrombosis and haemostasis, alternative therapy development including gene therapy and recombinant factor IX production for safer protein replacement therapy, and its regulation at the gene expression level. As one of the key factors in the blood coagulation cascade, factor IX will continue to serve as an invaluable model to provide fascinating insights into the intricate mechanism of blood coagulation and its regulation.

Acknowledgements—This work was in part supported by research grants from the National Institutes of Health (HL38644) and the Haemophilia Foundation of Michigan. We thank Mrs Nancy Remaklus for her help in preparing this manuscript.

References

1. Saito H. Normal haemostatic mechanisms. In: Ratnoff OD and Forbes CO, eds. *Disorders of Haemostasis*, 2nd ed. Philadelphia: WB Saunders, 1991: 18–47.
2. Davie EW, Fujikawa K, Kisiel W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* 1991; **30:** 10363–10370.
3. Naito K, Fujikawa K. Activation of human blood coagulation factor XI independence of factor XII. *J Biol Chem* 1991; **266:** 7353–7358.
4. Gailani D, Broze Jr GJ. Factor XI activation in a revised model of blood coagulation. *Science* 1991; **253:** 909–912.
5. Broze Jr GJ, Girard TJ, Novotny WF. Regulation of coagulation by a multivalent Kunitz-type inhibitor. *Biochemistry* 1990; **29:** 7539–7546.
6. Rapaport SI. The extrinsic pathway inhibitor: a regulation of tissue factor-dependent blood coagulation. *Thromb Haemostas* 1991; **66:** 6–15.
7. Bauer KA, Kass BL, ten Cate H, *et al.* Factor IX is activated *in vivo* by the tissue factor mechanism. *Blood* 1990; **76:** 731–736.
8. Kurachi K, Davie EW. Isolation and characterization of a cDNA coding for human factor IX. *Proc Natl Acad Sci USA* 1982; **79:** 6461–6465.
9. Yoshitake S, Schach BG, Foster DC, *et al.* Nucleotide sequence of the gene for human factor IX (antihaemophilic factor B). *Biochemistry* 1985; **24:** 3736–3750.
10. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* 1991; **266:** 19867–19870.
11. Pang CP, Crossley M, Brownlee GG. Comparative sequence analysis of mammalian factor IX promoters. *Nucleic Acids Res* 1990; **18:** 6731–6732.
12. Bergenhem NCH, Venta PJ, Hopkins PJ, *et al.* Mutation creates an open reading frame within the 5′ untranslated region of macaque erythrocyte carbonic anhydrase (CA) I mRNA that suppresses CA I expression and supports the scanning model for translation. *Proc Natl Acad Sci USA* 1992; **89:** 8798–8802.
13. Kawabata S, Davie EW. A microsomal endopeptidase from liver with substrate specificity for processing proproteins such as the vitamin K–dependent proteins of plasma. *J Biol Chem* 1992; **267:** 10331–10336.
14. Wise R, Barr PH, Wong PA, *et al.* Expression of a human proprotein processing enzyme: current cleavage of the von Willebrand factor precursor at a paired basic amino acid site. *Proc Natl Acad Sci USA* 1990; **87:** 9378–9382.
15. Wasley L-C, Rehemtulla A, Bristol JA, *et al.* PACE/Furin can process the vitamin K-dependent pro-factor IX precursor within the secretory pathway. *J Biol Chem* 1993; **268:** 8458–8465.
16. Meulien P, Balland A, Lepage P, *et al.* Increased biological activity of a recombinant factor IX variant carrying alanine at position +1. *Protein Eng* 1990; **3:** 629–633.
17. Cheung A, Suttie JW, Bernatowicz M. Vitamin K-dependent carboxylase: structural requirements for propeptide activation. *Biochim Biophys Acta* 1990; **1039:** 90.
18. Furie B, Furie BC. Molecular basis of γ-carboxylation: role of the propeptide in the vitamin K-dependent proteins. *Ann NY Acad Sci* 1991; **614:** 1–10.
19. Foster DC, Rudinski MS, Schach BG, *et al.* Propeptide of human protein C is necessary for γ-carboxylation. *Biochemistry* 1987; **26:** 7003–7011.
20. Diuguid DL, Rabiet MJ, Furie BC, *et al.* Molecular basis of haemophilia B: a defective enzyme due to an unprocessed propeptide is caused by a point mutation in the factor IX precursor. *Proc Natl Acad Sci USA* 1986; **83:** 5803–5807.
21. Bentley AK, Rees DJG, Rizza C, *et al.* Defective propeptide processing of blood clotting factor IX caused by mutation of arginine to glutamine. *Cell* 1986; **45:** 343–348.
22. Wu SM, Cheung WF, Frazier D, *et al.* Cloning and expression of the cDNA for human γ-glutamyl carboxylase. *Science* 1991; **254:** 1634–1636.
23. Rehemtulla A, Roth DA, Wasley L, *et al.* In vitro and in vivo functional characterization of bovine vitamin K-dependent γ-carboxylase expressed in Chinese hamster ovary cells. *Proc Natl Acad Sci USA* 1993; **90:** 4611–4615.
24. Kaufman RJ, Wasley LC, Furie, BC, *et al.* Expression, purification, and characterization of recombinant factor IX synthesized in Chinese hamster ovary cells. *J Biol Chem* 1986; **261:** 9622–9628.
25. Kurachi K. Recombinant antihaemophilic factors. In: Goldstein J, ed. *Biotechnology of Blood*, Boston: Butterworth-Heinemann, 1991: 177–195.
26. Balland A, Faure T, Carvallo D, *et al.* Characterization of two differently processed forms of human recombinant factor IX synthesized in CHO cells transformed with a polycistronic vector. *Eur J Biochem* 1988; **172:** 565–572.
27. Verweij CL, Hart M, Pannekoek H. Expression of variant von Willebrand factor (vWF) cDNA in heterologous cells: requirement of the pro-polypeptide in vWF multimer formation. *EMBO J* 1987; **6:** 2885–2890.
28. Wise RJ, Pittman DD, Handin RI, Kaufman RJ, Orkin SH. The propeptide of the von Willebrand factor independently mediates the assembly of von Willebrand multimers. *Cell* 1988; **52:** 229–236.
29. Sadler JE. von Willebrand factor. *J Biol Chem* 1991; **266:** 22777–22780.
30. Giannelli F, Green PM, High KA, *et al.* Haemophilia B: database of point mutations and short additions and deletions. *Nucleic Acids Res* 1992; **20:** 2027–2063.
31. Thompson AR. Molecular biology of the haemophilias. *Prog Haemostasis Thromb* 1991; **10:** 175–214.
32. Wang NS, Zhang M, Thompson AR, *et al.* Factor $IX_{Chongqing}$: a new mutation in the calcium binding domain of factor IX resulting in severe haemophilia B. *Thromb Haemostas* 1990; **63:** 24–26.
33. Stenflo J. Structure-function relationships of epidermal growth factor modules in vitamin K-dependent clotting factors. *Blood* 1991; **78:** 1637–1651.
34. Astermark J, Bjork I, Ohlin A-K, *et al.* Structural requirements for Ca^{2+} binding to the γ-carboxyglutamic acid and epidermal growth factor-like regions of factor IX. *J Biol Chem* 1991; **266:** 2430–2437.
35. Tulinsky A. The structure of domains of blood proteins. *Thromb Haemostas* 1991; **66:** 16–30.
36. Soriano-Garcia M, Padmanabhan K, de Vos AM, *et al.* The Ca^{2+} ion and membrane binding structure of the Gla domain of Ca-prothrombin fragment 1. *Biochemistry* 1992; **31:** 2554–2566.

37. Wolberg AS, Cheung W-F, Stafford DW. Characterization of the functional roles of γ-carboxylglutamic acid (Gla) residues 20 and 21 in human factor IX. *Thromb Haemostas* 1993; **69:** 614.
38. Cheung W-F, Hamaguchi N, Smith KJ, *et al.* The binding of human factor IX to endothelial cells is mediated by residues 3–11. *J Biol Chem* 1992; **267:** 20529–20533.
39. Huang LH, Ke X-H, Sweeney W, *et al.* Calcium binding and putative activity of the epidermal growth factor domain of blood coagulation factor IX. *Biochem Biophys Res Commun* 1989; **160:** 133–139.
40. Selander M, Persson E, Stenflo J, *et al.* ^{1}H NMR assignment and secondary structure of the Ca^{2+}-free form of the amino-terminal epidermal growth factor-like domain in coagulation factor X. *Biochemistry* 1990; **29:** 8111–8118.
41. Handford PA, Baron M, Mayhew M, *et al.* The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO J* 1990; **9:** 475–480.
42. Fernlund P, Stenflo J. β-hydroxyaspartic acid in vitamin K-dependent proteins. *J Biol Chem* 1983 **258:** 12509–12512.
43. Hase S, Nishimuru H, Kawabata S-I, *et al.* The structure of $(Xylose)_2$Glucose-*O*-Serine 53 found in the first epidermal growth factor-like domain of bovine blood clotting factor IX. *J Biol Chem* 1990; **265:** 1858–1861.
44. Derian CK, VanDusen W, Przysiecki CT, *et al.* Inhibitors of 2-ketoglutarate dependent dioxygenases block aspartyl β-hydroxylation of recombinant human factor IX in several mammalian expression systems. *J Biol Chem* 1989; **264:** 6615–6618.
45. Johnson AE, Esmon NL, Lane TM, *et al.* Structural changes required for activation of protein C are induced by Ca^{2+} binding to a high affinity site that does not contain γ-carboxyglutamic acid. *J Biol Chem* 1983; **258:** 5554.
46. Davis LM, McGraw RA, Ware JL, *et al.* Factor $IX_{Alabama}$: a point mutation in a clotting protein results in haemophilia B. *Blood* 1987; **69:** 140–143.
47. Lozier JN, Monroe DM, Stanfield-Oakley S, *et al.* Factor $IX_{New\ London}$: Substitution of Proline for glutamine at position 50 causes severe haemophilia B. *Blood* 1990; **75:** 1097–1104.
48. Spitzer SG, Kuppuswamy MN, Saini R, *et al.* Factor $IX_{Hollywood}$: Substitution of Pro55 by Ala in the first epidermal growth factor-like domain. *Blood* 1990; **76:** 1530–1537.
49. Lin SW, Smith KJ, Welsch D, *et al.* Expression and characterization of human factor IX and factor IX-factor X chimeras in mouse C127 cells. *J Biol Chem* 1990; **265:** 144–150.
50. Miyata T, Nishimura H, Suehiro K, *et al.* The 2nd EGF-like domain of factor IX is required for interaction with factor VIII. *Thromb Haemostas* 1991; **65:** 817.
51. Hertzberg MS, Ben-Tal O, Furie B, *et al:* Construction, expression, and characterization of a chimera of factor IX and factor X: the role of the second epidermal growth factor domain and serine protease domain in factor Va binding. *J Biol Chem* 1992; **267:** 14759–14766.
52. Noyes CM, Griffith MJ, Roberts HR, *et al.* Identification of the molecular defect in factor $IX_{Chapel\ Hill}$: substitution of histidine for arginine at position 145. *Proc Natl Acad Sci USA* 1983; **80:** 4200–4202.
53. Liddell MB, Peake IR, Taylor SAM, *et al.* Factor $IX_{Cardiff}$: a variant factor IX protein that shows abnormal activation is caused by an arginine to cysteine substitution at position 145. *Br J Haematol* 1989; **72:** 556–560.
54. Di Scipio RG, Kurachi K, Davie EW. Activation of human factor IX (Christmas factor). *J Clin Invest* 1978; **61:** 1528–1538.
55. Bajaj SP, Spitzer SG, Welsh WJ, *et al.* Experimental and theoretical evidence supporting the role of Gly363 in blood coagulation factor IXa (Gly193 in chymotrypsin) for proper activation of the proenzyme. *J Biol Chem* 1990; **265:** 2956–2961.
56. Usharani P, Warn-Cramer BJ, Kasper CK, *et al.* Characterization of three abnormal factor IX variants (Bm Lake Elsinore, Long Beach, and Los Angeles) of Haemophilia-B. *J Clin Invest* 1985; **75:** 76–83.
57. Bottema CDK, Ketterling RP, Li S, *et al.* Missense mutations and evolutionary conservation of amino acids: evidence that many of the amino acids in factor IX function as "spacer" elements. *Am J Hum Genet* 1991; **49:** 820–838.
58. Bajaj SP, Sabarwal AK, Gorka J, *et al.* Antibody-probed conformational transitions in the protease domain of factor IX upon calcium binding and zymogen activation: putative high affinity calcium binding site in the protease domain. *Proc Natl Acad Sci USA* 1992; **89:** 152–156.
59. Sinha U, Wolf DL. Carbohydrate residues modulate the activation of coagulation factor X. *J Biol Chem* 1993; **268:** 3048–3051.
60. Choo KH, Gould KG, Rees DJG, *et al.* Molecular cloning of the gene for human anti-haemophilic factor IX. *Nature* 1982; **299:** 178–180.
61. Chance PF, Dyer KA, Kurachi K, *et al.* Regional localization of human factor IX gene by molecular hybridization. *Hum Genet* 1983; **65:** 207–208.
62. Purrello M, Alhadeff B, Esposito D, *et al.* The human genes for haemophilia A and haemophilia B flank the X chromosome fragile site at Xq27.3. *EMBO J* 1985; **4:** 725–729.
63. Camerino G, Grzeschik KH, Jaye M, *et al.* Regional localization on the human X chromosome and polymorphism of the coagulation factor IX gene (haemophilia B locus). *Proc Natl Acad Sci USA* 1984; **81:** 498–502.
64. Bottema CD, Bottema MJ, Ketterling RP, *et al.* Why does the human factor IX gene have a G + C content of 40%? *Am J Hum Genet* 1991; **49:** 839–850.
65. Foster DJ, Yoshitake S, Davie EW. The molecular sequence of the gene for human protein C. *Proc Natl Acad Sci USA* 1985; **82:** 4673–4677.
66. Schmidel DK, Tatro AV, Phelps LG, *et al.* Organization of the human protein S genes. *Biochemistry* 1990; **29:** 7845–7852.
67. Degen SJF, Davie EW. Nucleotide sequence of the gene for human prothrombin. *Biochemistry* 1987; **26:** 6165–6177.
68. Hojrup P, Jensen MS, Petersen TE. Amino acid sequence of bovine protein Z: a vitamin K-dependent serine protease homology. *FEBS Lett* 1985; **184:** 333–338.
69. Seijima H, Hayashi T, Deyashiki Y, *et al.* Primary structure of vitamin K-dependent human protein Z. *Biochem Biophys Res Commun* 1990; **171:** 661–666.

70. Leytus SP, Foster DC, Kurachi K, *et al.* Gene for human factor X: a blood coagulation factor whose gene organization is essentially identical with that of factor IX and protein C. *Biochemistry* 1986; **25:** 5098–5102.
71. O'Hara PJ, Grant FJ, Haldeman BA, *et al.* Nucleotide sequence of the gene coding for human factor VII, a vitamin K-dependent protein participating in blood coagulation. *Proc Natl Acad Sci USA* 1987; **84:** 5158–5162.
72. Hay CW, Robertson KA, Young SN, *et al.* Use of Bam HI polymorphism in the factor IX gene for the determination of haemophilia B carrier status. *Blood* 1986; **67:** 1508.
73. Winship PR, Anson DS, Rizza CR, *et al.* Carrier detection in haemophilia B using two further intragenic restriction fragment length polymorphisms. *Nucleic Acids Res* 1984; **12:** 8861–8872.
74. Camerino G, Grzeschik KH, Jaye M, *et al.* Regional localization on the human X-chromosome and polymorphism of the coagulation factor IX gene (haemophilia B locus). *Proc Natl Acad Sci USA* 1984; **81:** 498–502.
75. McGraw RA, Davis LM, Noyes CM, *et al.* Evidence for a prevalent dimorphism in the activation peptide of human coagulation factor IX. *Proc. Natl Acad Sci USA* 1985; **82:** 2847–2851.
76. Tanimoto M, Hamaguchi M, Matsushita T, *et al.* A new marker (F9–192) of the factor IX gene detected by denaturing gel electrophoresis. *Blood* 1989; **74:** 251a.
77. Driscoll MC, Dispenzieri A, Tobias E, *et al.* A second Bam HI DNA polymorphism and haplotype association in the factor IX gene. *Blood* 1988; **72:** 61–65.
78. Camerino G, Oberle I, Drayana D, *et al.* A new Msp I restriction fragment length polymorphism in the haemophilia B locus. *Hum Genet* 1985; **71:** 79–81.
79. Winship PR, Rees DJG, Alkan M. Detection of polymorphisms at cytosine phosphoguanadine dinucleotides and diagnosis of haemophilia B carriers. *Lancet* 1989; **25:** 631–634.
80. Reiner AP, Thompson AR. An Hha I polymorphism is present in factor IX genes of Asian subjects. *Hum Genet* 1990; **86:** 87–88.
81. Mulligan L, Holden JJA, White BN. A DNA marker closely linked to the factor IX (haemophilia B) gene. *Hum Genet* 1987; **75:** 381–383.
82. Hofker MH, Bergen AAR, Skraastad MI, *et al.* Efficient isolation of X chromosome-specific single-copy probes from a cosmid library of a human X/hamster hybrid-cell line: mapping of new probes close to the locus for X-linked mental retardation. *Am J Hum Genet* 1987; **40:** 312–328.
83. Kurachi K, Chen S-H. Molecular genetic analyses of congenital disorders of coagulation factors. In: Bern MM, Frigoletto F, eds. *Hematologic Disorders in Maternal–Fetal Medicine.* New York: Wiley-Liss, Inc., 1989: 449–466.
84. Sarkar G, Paynton C, Sommer SS. Segments containing alternating purine and pyrimidine dinucleotides: patterns of polymorphism in humans and prevalence throughout phylogeny. *Nucleic Acids Res* 1991; **19:** 631–636.
85. Kaplan DJ, Duncan CH. Novel short interspersed repeat in human DNA. *Nucleic Acids Res* 1990; **18:** 192.
86. Chen S-H, Scott CR. Recombination between two 14-bp homologous sequences as the mechanism for gene deletion in factor $IX_{Seattle\ 1}$. *Am J Hum Genet* 1990; **47:** 1020–1022.
87. Vidaud M, Vidaud D, Siguret V, *et al.* Mutational insertion of an Alu sequence causes haemophilia B. *Am J Hum Genet* 1989; **45:** A226.
88. Chen S-H, Scott CR, Edson R, *et al.* An insertion within the factor IX gene for $haemophilia_{El\ Salvadore}$. *Am J Hum Genet* 1988; **42:** 581–584.
89. Kazazian HH, Wong C, Youssoufian H, *et al.* Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 1988; **332:** 164–166.
90. Youssoufian H, Kazazian HH, Patel A, *et al.* Mild haemophilia A associated with a cryptic donor splice site mutation in intron 4 of the factor VIII gene. *Genomics* 1988; **2:** 32–36.
91. Toole JJ, Pittman DD, Orr EC, *et al.* A large region (X95 kDa) of human factor VIII is dispensable for *in vitro* procoagulant activity. *Proc Natl Acad Sci USA* 1986; **83:** 5939–5942.
92. Pittman DD, Alderman EM, Tomkinson KN, *et al.* Biochemical, immunological, and *in vivo* functional characterization of B-domain-deleted factor VIII. *Blood* 1993; **81:** 2925–2935.
93. Koeberl DD, Bottema CD, Buerstedde J-M, *et al.* Functionally important regions of the factor IX gene have a low rate of polymorphism and a high rate of mutations in the dinucleotide CpG. *Am J Hum Genet* 1989; **45:** 448–457.
94. Sommer SS. Assessing the underlying pattern of human germline mutations: lesions from the factor IX gene. *FASEB J* 1992; **6:** 2767–2774.
95. Higuchi M, Kazazian HH, Jr, Kasch L, *et al.* Molecular characterization of severe haemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene. *Proc Natl Acad Sci USA* 1991; **88:** 7405–7409.
96. Ketterling RP, Bottema CD, Koeberl DD, *et al.* T296IM, a common mutation causing mild haemophilia B in the Amish and others: founder effect, variability in factor IX activity assays and rapid carrier detection. *Hum Genet* 1991; **87:** 333–337.
97. Taylor SAM, Deugan KV, Lillicrap DP. Somatic mosaicism and female-40-female transmission in a kindred with haemophilia B. *Proc Natl Acad Sci USA* 1991; **88:** 39–42.
98. Nisen P, Stamberg J, Ehrenpreis R, *et al.* The molecular basis of severe haemophilia B in a girl. *N Engl J Med* 1986; **315:** 1139–1142.
99. Salier J-P, Hirosawa S, Kurachi K. Functional characterization of the 5′-regulatory region of human factor IX gene. *J Biol Chem* 1990; **265:** 7062–7068.
100. Sarkar G, Sommer SS. Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science* 1989; **244:** 331–334.
101. Hassan HJ, Leonardi A, Chelucci C, *et al.* Blood coagulation factors in human embryonic-fetal development: preferential expression of the FVII/tissue factor pathway. *Blood* 1990; **76:** 1158–1164.
102. Kisker CT, Robillard JE, Clarke WR. Development of blood coagulation—a fetal lamb model. *Pediatr Res* 1981; **15:** 1045–1050.

103. Yao S-N, DeSilva AH, Kurachi S, *et al.* Characterization of a mouse factor IX cDNA and developmental regulation of the factor IX gene expression in liver. *Thromb Haemostas* 1991; **65:** 52–58.
104. Andrew M, Paes B, Milner R, *et al.* Development of the human coagulation system in the full-term infant. *Blood* 1987; **70:** 165–172.
105. Andrew M, Paes B, Milner R, *et al.* Development of the human coagulation system in the healthy premature infant. *Blood* 1988; **72:** 1651–1657.
106. Thompson AR. Factor IX antigen by radioimmunoassay: abnormal factor IX protein in patients on warfarin and with haemophilia B. *J Clin Invest* 1977; **59:** 900–910.
107. Lane PA, Hathaway WE. Vitamin K in infancy. *J Pediatr* 1985; **106:** 351–359.
108. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells of sequence–specific DNA binding proteins. *Science* 1989; **245:** 371–378.
109. Anson DS, Choo KH, Rees DJG, *et al.* The gene structure of human anti-haemophilic factor IX. *EMBO J* 1984; **3:** 1053–1060.
110. Reijnen M, Bertina RM, Reitsma PH. Localization of transcription initiation sites in the human coagulation factor IX gene. *FEBS Lett* 1990; **270:** 207–210.
111. Kurachi S, Furukawa M, Salier J-P, *et al.* Regulatory mechanism of human factor IX gene: Protein bindings at the Leyden-specific region. Submitted for publication.
112. Kurachi K, Furukawa M, Yao S-N, *et al.* Regulation of the factor IX gene. *Thromb Haemostas* 1991; **65:** 940.
113. Evans JP, Watzke HH, Ware JL, *et al.* Molecular cloning of a cDNA encoding canine factor IX. *Blood* 1989; **74:** 207–212.
114. Crossley M, Brownlee GG. Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. *Nature* 1990; **345:** 444–446.
115. Crossley M, Ludwig M, Stowell KM, *et al.* Recovery from haemophilia B Leyden: an androgen-responsive element in the factor IX promoter. *Science* 1992; **257:** 377–379.
116. Mendel DB, Crabtree GR. HNF-1, a member of a novel class of dimerizing homeodomain proteins. *J Biol Chem* 1991; **266:** 677–680.
117. Courtois G, Baumhueter S, Crabtree GR. Purified hepatocyte nuclear factor 1 interacts with a family of hepatocyte-specific promoters. *Proc Natl Acad Sci USA* 1988; **85:** 7937–7941.
118. Saffer JD, Thurston SJ. A negative regulatory element with properties similar to those of enhancers is contained within an Alu sequence. *Mol Cell Biol* 1989; **9:** 355–364.
119. Jallat S, Perraud F, Dalemans W, *et al.* Characterization of recombinant human factor IX expressed in transgenic mice and in derived transimmortalized hepatic cell lines. *EMBO J* 1990; **9:** 3295–3301.
120. Briet E, Bertina RM, van Tilburg NH, *et al.* Haemophilia B Leyden. *N Engl J Med* 1982; **306:** 788–790.
121. Mandalaki T, Louizou C, Dimitriadou C, *et al.* Haemophilia B Leyden in Greece. *Thromb Haemostas* 1986; **56:** 340–342.
122. Hirosawa S, Fahner JB, Salier J-P, *et al.* Structural and functional basis of the developmental regulation of human coagulation factor IX gene: factor IX Leyden. *Proc Natl Acad Sci USA* 1990; **87:** 4421–4425.
123. Reitsma PH, Bertina RM, van Amstel JKP, *et al.* The putative factor IX gene promoter in haemophilia B Leyden. *Blood* 1988; **72:** 1074–1076.
124. Attree O, Vidaud M, Vidaud D, *et al.* New strategies for rapid detection and characterization of mutations in haemophilia B. *Thromb Haemostas* 1989; **62:** 8.
125. Picketts DJ, Lillicrap DP, Mueller CR. An A to T transversion at nucleotide −5 in the factor IX promoter disrupts a transcription factor binding site between nucleotides −1 and −15. *Blood* 1991; **78** (suppl 1): 180a.
126. Friedenberg DL, Black B. Altered developmental control of the factor IX gene: a new T to A mutation at position +6 of the FIX gene resulting in haemophilia B Leyden. *Thromb Haemostas* 1991; **65:** 964.
127. Royle G, Van de Water NS, Berry E, *et al.* Haemophila B Leyden arising *de novo* by point mutation in the putative factor IX promoter region. *Br J Haematol* 1991; **77:** 191–194.
128. Rietsma PH, Mandalaki T, Kasper CK, *et al.* Two novel point mutations correlate with an altered developmental expression of blood coagulation factor IX (Haemophilia B Leyden phenotype). *Blood* 1989; **73:** 743–746.
129. Reijnen, MJ, Sladek FM, Bertina RM, *et al.* Disruption of a binding site for hepatocyte nuclear factor 4 results in haemophilia B Leyden. *Proc Natl Acad Sci USA* 1992; **89:** 6300–6303.
130. Kurachi S, Furukawa M, Kurachi K. Regulatory mechanism of factor IX gene which involves dynamic DNA-protein and protein-protein interactions: a possible mechanism of amelioration of factor IX-Leyden gene expression. *Thromb Haemostas* 1993; **69:** 847.
131. Picketts DJ, Lillicrap DP, Mueller CR. Synergy between transcription factors DBP and C/EBP compensates for a haemophilia B Leyden factor IX mutation. *Nature Genet* 1993; **3:** 175–179.
132. Berkhout B, Silverman RH, Heang KT. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* 1989; **59:** 273–282.
133. Laspia MF, Rice AP, Mathews MB. HIV-1 Tat protein increases transcriptional initiations and stabilizes elongation. *Cell* 1989; **59:** 283–292.
134. Theil EC. Regulation of ferritin and transferrin receptor mRNAs. *J Biol Chem* 1990; **265:** 4771–4774.
135. Kitchens CS. Haemophilia and von Willebrand's disease. In: Rabel RE, ed. *Conn's Current Therapy 1991.* Philadelphia: WB Saunders, 1991: 331–337.
136. Spero JA. Complications of the treatment of haemostatic disorders. In: Ratnoff OD, Forbes CD, eds. *Disorders of Haemostasis.* Philadelphia: WB Saunders, 1991: 555–575.
137. Anderson WF. Human gene therapy. *Science* 1992; **256:** 808–813.
138. Miller AD. Human gene therapy comes of age. *Nature* 1992; **357:** 455–460.
139. St. Louis D, Verma IM. An alternative approach to somatic cell gene therapy. *Proc Natl Acad Sci USA* 1988; **85:** 3150–3154.
140. Palmer TD, Thompson AR, Miller AD. Production of human factor IX in animals by genetically modified skin fibroblasts: potential therapy for haemophilia B. *Blood* 1989; **73:** 438–445.

141. Axelrod JH, Read MS, Brinkhous KM, Verma IM. Phenotypic correction of factor IX deficiency in skin fibroblasts of haemophilic dogs. *Proc Natl Acad Sci USA* 1990; **87:** 5173–5177.
142. Yao S-N, Wilson JM, Nabel EG, *et al.* Expression of human factor IX in rat capillary endothelial cells: Toward somatic gene therapy for haemophilia B. *Proc Natl Acad Sci USA* 1991; **88:** 8101–8105.
143. Armentano D, Thompson AR, Darlington G, *et al.* Expression of human factor IX in rabbit hepatocytes by retrovirus-mediated gene transfer: potential for gene therapy of haemophilia B. *Proc Natl Acad Sci USA* 1990; **87:** 6141–6145.
144. Yao S-N, Kurachi K. Expression of human factor IX in mice after injection of genetically modified myoblasts. *Proc Natl Acad Sci USA* 1992; **89:** 3357–3361.
145. Gerrard AJ, Hudson DL, Brownlee GG, *et al.* Towards gene therapy for haemophilia B using primary human keratinocytes. *Nature Genet* 1993; **3:** 180–183.
146. Yao S-N, Kurachi K. A simple treatment of serum for precise determination of recombinant factor IX in the culture media. *BioTechniques* 1991; **12:** 525–526.
147. Scharfmann R, Axelrod JH, Verma IM. Long-term *in vivo* expresssion of retrovirus-mediated gene transfer in mouse fibroblast implants. *Proc Natl Acad Sci USA* 1991; **88:** 4626–4630.
148. Hseuh JL, Lu DR, Zhou JM, *et al.* Gene therapy for haemophilia B using fibroblasts. Part I. *Sci China* (Ser B), 1993; **23:** 53–60.
149. Yao S-N, Wilson JM, Nabel EG, *et al.* Expression of human factor IX in rat capillary endothelial cells: Toward somatic gene therapy for haemophilia B. *Proc Natl Acad Sci USA* 1991; **88:** 8101–8105.
150. Ponder KP, Gupta S, Feland F, *et al.* Mouse hepatocytes migrate to liver parenchyma and function indefinitely after intrasplenic transplantation. *Proc Natl Acad Sci USA* 1991; **88:** 1217–1221.
151. Kay, MA, Baley P, Rothenberg S, *et al.* Expression of human α_1-antitrypsin in dogs after autologous transplantation of retroviral transduced hepatocytes. *Proc Natl Acad Sci USA* 1992; **89:** 89–93.
152. Ferkol T, Lindberg GL, Chen J, *et al.* Regulation of the phosphenolpyruvate carboxykinase/human factor IX gene introduced into the livers of adult rats by receptor-mediated gene transfer. *FASEB J* 1993; **7:** 1081–1091.
153. Kay MA, Rothenberg S, Lande CN, *et al. In vivo* gene therapy of haemophilia B: sustained partial correction in factor IX-deficient dogs. *Science* 1993; **263:** 117–119.

Am. J. Hum. Genet. 49:820–838, 1991

Exhibit C

Missense Mutations and Evolutionary Conservation of Amino Acids: Evidence That Many of the Amino Acids in Factor IX Function as "Spacer" Elements[1]

Cynthia D. K. Bottema,* Rhett P. Ketterling,* Setsuko Ii,* Hong-Sup Yoon,* John A. Phillips III,† and Steve S. Sommer*

*Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, MN; and †Division of Genetics, Department of Pediatrics, Vanderbilt University School of Medicine, Nashville

Summary

We report 31 point mutations in the factor IX gene and explore the relationship between the level of evolutionary conservation of an amino acid and the probability of a mutation causing hemophilia B. From our total sample of 125 hemophiliacs and from those reported by others, we identify 95 independent missense mutations, 94 of which occur at amino acids that are evolutionarily conserved in the available mammalian factor IX sequences. The likelihood of a missense mutation causing hemophilia B depends on whether the residue is also conserved in the factor IX–related proteases: factor VII, factor X, and protein C. Most of the possible missense mutations in generically conserved residues (i.e., those conserved in factor IX and in all the related proteases) should cause disease. In contrast, missense mutations in factor IX–specific residues (i.e., those conserved in human, cow, dog, and mouse factor IX but *not* in the related proteases) are sixfold less likely to cause disease. Missense mutations at nonconserved residues are 33-fold less likely to cause disease. At least three models are compatible with these observations. A comparison of sequence alignments from four and nine species of factor IX and an examination of the missense mutations occurring at CpG residues suggest a model in which most residues fall on opposite ends of a spectrum. In about 40% of residues, virtually any missense mutation in a minority of the residues will cause disease, while virtually *no* missense mutations will cause disease in most of the remaining residues. Thus, many of the residues in factor IX are spacers; that is, the main chains are presumably necessary to keep other amino acid interactions in register, but the nature of the side chain is unimportant.

Introduction

Factor IX is a coagulation serine protease zymogen with eight functional domains, including (1) a signal peptide, (2) a pro-peptide which is necessary for the γ-carboxylation of the mature protein, (3) a gla domain with 12 γ-carboxyglutamic (gla) residues which bind four to six molecules of calcium, (4) a short aromatic amino acid stack, (5) a first epidermal growth factor domain which contains a high-affinity calcium-binding site, (6) a second epidermal growth factor domain of unknown function, (7) an activation peptide which is removed during proteolysis by factors VII or XI, and (8) a catalytic domain which activates factor X (reviewed in Furie and Furie 1988). Factor IX, factor VII, factor X, and protein C are closely related coagulation serine proteases that have the same eight functional domains and similar gene structure (Furie and Furie 1988).

Since the factor IX gene is located on the X chromosome, a mutation that disrupts function affects any male who receives that allele. Many mutations in the factor IX gene causing hemophilia B have been described (reviewed in Giannelli et al. 1990). The mutation rate is dramatically enhanced at CpG dinucleotides (Koeberl et al. 1989; Green et al. 1990) but not at any other dinucleotides (Bottema et al. 1991).

Received April 18, 1991; final revision received June 17, 1991.

Address for correspondence and reprints: Dr. Steve S. Sommer, Department of Biochemistry and Molecular Biology, Mayo Clinic Foundation, Rochester, MN 55905.

1. This paper is dedicated to Harold and Hadassah Sommer.

0002-9297/91/4904-0015$02.00

Herein we present 31 new families with point mutations and analyze the relationship between evolutionary conservation of amino acids and missense mutations which cause hemophilia B in humans.

Methods

Sequencing

DNA was extracted from blood collected in ACD solution B or solution A as previously described (Gustafson et al. 1987). Regions of likely functional significance were sequenced by genomic *a*mplification *w*ith *t*ranscript *s*equencing (GAWTS) (Stoflet et al. 1988) as described by Sommer et al. (1990). GAWTS is a method of direct sequencing that involves (1) PCR amplification of the segment of interest, where at least one of the PCR primers has an attached phage promoter sequence, (2) transcription of the amplified segment with the phage RNA polymerase to produce a single-stranded RNA molecule, and (3) sequencing of the RNA template with reverse transcriptase. The following bases were sequenced (numbering system corresponds to that of Yoshitake et al. [1985]): region A, −106 to 139; region B/C, 6720 to 6265; region D, 10544 to 10315; region E, 17847 to 17601; region F, 20577 to 20334; region G, 30183 to 29978; and region H, 31411 to 30764. The poly A addition region was not sequenced. The order of the numbers in each region indicates the direction of sequencing. In all, at least 2.2 kb was sequenced from each hemophiliac.

Haplotype Analysis

The following polymorphisms in the factor IX gene were examined: *Hin*fI (intron a) (Winship et al. 1984), *Xmn*I (intron c) (Winship et al. 1984), *Taq*I (intron d) (Camerino et al. 1984), and *Hha*I (3′ of gene) (Winship et al. 1989). From these four polymorphisms, eight common haplotypes were defined, with frequencies of 2%–19% in the normal Caucasian population (Ketterling et al., in press).

DNA segments containing the *Taq*I and the *Xmn*I restriction sites were amplified by PCR and digested with the appropriate restriction enzyme as described elsewhere (Koeberl et al. 1990). The products were gel electrophoresed, and the presence (+) or the absence (−) of the restriction site was determined. For the *Hin*fI (also known as *Dde*I) polymorphism, the DNA was amplified by PCR, and the presence (+) or absence (−) of the 50-bp insert was determined by gel electrophoresis. The *Hha*I polymorphism was determined by amplifying 500 ng genomic DNA by PCR using 0.1 μM of the previously described oligonucleotides H1 and H2 (Winship et al. 1989) and 1.5 mM $MgCl_2$ in 50-μl reactions. The PCR products were digested with *Hha*I, and the presence (+) or absence (−) of the restriction site was determined by gel electrophoresis.

Levels of Amino Acid Conservation in Factor IX

Four classes of residues can be defined from the available factor IX sequences and from the sequences of both human and bovine factor VII, factor X, and protein C (fig. 1 and Appendixes A and B). A residue is "generic" if it is identical in all species of factor IX and is also identical in the three related blood coagulation serine proteases: factor VII, factor X, and protein C. The residue is "factor IX specific" if it is identical in all species of factor IX but *not* identical in any of the three related proteases. The residue is "partially generic" if it is identical in all species of factor IX and is identical in one or two of the three related proteases. If a residue is conservatively substituted in the species of factor IX (i.e., is S/T, S/A, Y/F, R/K, I/L/V, N/D, D/E, Q/N, and E/Q), the above definitions are modified to also allow the conservative substitution in the three related proteases. If a residue is nonconservatively substituted in any of the species of factor IX, it is classified as "nonconserved."

The above definitions differ from a previous classification (Koeberl et al. 1990; Sarkar et al. 1990) in that the sequence alignment utilizes the complete sequence of dog factor IX (Evans et al., 1989) and mouse factor IX (Wu et al. 1990) and bovine factor VII (Takeya et al. 1988) (Appendixes A and B). Most important, the presently defined factor IX–specific residues and most of the partially generic residues were previously combined into one class.

A detailed protocol was used for assigning residues to each class (see Identity subsection). Our conclusions will remain the same despite certain revisions in the classification protocol (see Conservative substitutions subsection).

1. *Identity.*—Those amino acids identical in the mammalian factor IX sequences were compared with the homologous residues in the related coagulation serine proteases. The amino acid was assigned to a class on the basis of extent of identity with these proteases. For a residue to be considered conserved in a given related protease, both the human and bovine residues needed to be identical with the corresponding factor IX amino acid.

2. *Conservative substitutions* — Each amino acid which was not identical in the factor IX sequences was considered to be nonconserved unless the substitutions were highly conservative. These highly conservative substitutions were S/T, S/A, E/D, D/N, Q/E, Q/N, F/Y, K/R, and I/L/V (see fig. 1 legend for the single-letter amino acid code). Conservative substitutions were defined rather stringently in that, with one exception (I/L/V), only two residues constitute each conservation group. As examples, D (aspartate) and N (asparagine) are in one group because they are related polar residues of approximately the same volume. D and E (glutamate) are in another group because they both have a negative charge. Thus, at any given D, the conservative substitutions were limited to *either* charge or size. Therefore, the presence of D and E in the factor IX sequences is classified as a conservative substitution. While the presence of D, E, and N is classified as nonconservative. (The alternative possibility of allowing *any* combination of either (*a*) D, E, Q, and N or (*b*) S, T, and A converts only four nonconserved residues to the conserved class and does not alter any of the conclusions). The conservatively substituted factor IX amino acids were then compared with the related serine proteases and were assigned to one of the above classes (i.e., generic, partially generic, or factor IX specific).

In practice, almost all of the generic, partially generic, and factor IX–specific residues are identical rather than conservatively substituted. Conservative substitutions in factor IX occur in only 8% of the 364 conserved amino acids. If conservative substitutions are not allowed, the number of residues involved in hemophilia B remains virtually unchanged (182 amino acids without substitutions vs. 176 amino acids with conservative substitutions [table 3]). However, the number of mutations in nonconserved residues increases from one to seven if conservative substitutions are not allowed. Mutations were observed at one N/D site, four I/L/V sites, one T/S conservative site (see fig. 1).

3. *Mutations at CpG versus non-CpG sites.* — Since mutations at the dinucleotide CpG occur more frequently than those at other sites (Koeberl et al. 1989; Green et al. 1990), the mutations at CpG and non-CpG nucleotides were analyzed separately (see tables 3 and 5). Residues at CpG dinucleotides were assigned to the non-CpG or CpG categories. This assignment was based on the fraction of possible mutations at CpG and non-CpG nucleotides in each residue (table 2). As examples, all of the arginine residues with codons of CGX were assigned to the CpG dinucleotide group. However, a glycine residue preceded by a residue ending in C (XXC GGX) had the first G assigned to the CpG group and the second G assigned to the non-CpG group. Although one-third of all independent mutations occur at CpG, the rarity of this dinucleotide in the factor IX coding sequence stipulates that the CpG group accounts for less than 3% of all the possible kinds of missense mutations.

Results

Mutations

Point mutations were delineated in 31 families with hemophilia B by direct genomic sequencing of the regions of likely functional significance, which include the coding region, the splice junctions, the putative promoter, the 5′ untranslated region, and a small part of the 3′ untranslated region (Koeberl et al. 1989). In total, 66 kb of sequence were obtained. While the majority of the hemophiliacs are Caucasians of northern-European descent, two (HB101 and HB102) are Hispanic and one (HB109) is Japanese.

Of the 31 point mutations, two affect splice junctions and six produce nonsense mutations, but the great majority (23) are missense mutations (table 1). Only one sequence change was found in each individual. The splice junction mutations which disrupt known consensus sequence, as well as the nonsense mutations which result in truncated protein products, are clearly causative mutations. In addition, we conclude that the missense mutations are also all causative because (1) they are the only sequence change found in the regions of likely functional significance, (2) polymorphisms in these regions are rare (Koeberl et al. 1989), (3) these changes are not present in normal individuals or as second site changes in other hemophiliacs, and (4) the missense mutations (except for one) are all at evolutionarily conserved residues (fig. 1).

Twenty-two of these mutations have not been previously described. Two mutations (serine^{94} and lysine^{412}) occur at residues specifically conserved in factor IX but not in the related proteases (fig. 1). Thus, these mutated residues may be important for factor IX–specific interactions such as binding to factors VIII or VII. Asparagine^{147}→isoleucine (HB108) is the first reported missense mutation in a hemophiliac to occur in a nonconserved residue. Tryptophan^{407}→arginine represents the first non-CpG site at which two patients (HB20 and HB92) have the same mutation in a differ-

Table 1

Single Base Mutations in Factor IX

Family	Factor IX Coagulant Reported (%)[d]	Nucleotide Change	Nucleotide Number	Structural Change	Domain	CpG Mutation[a]	Conservation Class[b]	Previous Report
HB64	<1	G→A	117	V[-17]→I	pro	No	P	
HB90	4	C→T	6364	R[-4]→W	pro	Yes	P	Giannelli et al. 1990
HB116	<1	G→A	6428	C[18]→Y	gla	No	G	
HB97	<1	G→T	10406	E[52]→TAG	EGF 1	No		
HB106	1	A→T	17697	R[94]→S	EGF 2	No	S	
HB111	<1[c]	T→C	17743	S[109]→P	EGF 2	No	P	
HB88	2	G→A	17786	C[124]→Y	EGF 2	No	G	
HB68	1	G→A	17797	V[128]→M	EGF 2	No	P	
HB115	<1	G→A	20375	C[133]→Y	EGF 2	No	G	
HB120	11	G→A	20414	R[145]→H	Activation peptide	Yes	P	Giannelli et al. 1990, Koeberl et al. 1990
HB102	<1	G→C	30038	Intron f, −1	Splice acceptor	No		
HB65	15	T→G	30101	I[216]→M	Catalytic	No	G	
HB114	<1	G→A	30821	Intron g, −1	Splice acceptor	No		Giannelli et al. 1990
HB98 and HB104	6 and 15	G→A	30864	R[238]→Q	Catalytic	Yes	P	Giannelli et al. 1990, Koeberl et al. 1990
HB91	<1[c]	C→T	30875	R[242]→TGA	Catalytic	Yes		Giannelli et al. 1988
HB96	<1	T→C	30930	I[260]→T	Catalytic	No	P	
HB101	8	T→C	30945	I[265]→P	Catalytic	No	G	Giannelli et al. 1990
HB109	<1[c]	T→G	30985	I[278]→M	Catalytic	No	P	
HB100	<1	G→T	31001	E[294]→TAA	Catalytic	No		
HB82	4	C→T	31091	Q[321]→TAG	Catalytic	No		
HB122	4	C→G	31096	Y[323]→TAG	Catalytic	No		
HB105	4	C→T	31118	R[333]→TGA	Catalytic	Yes		Giannelli et al. 1990; Koeberl et al. 1990
HB110	<1	G→T	31118	R[333]→L	Catalytic	Yes	P	
HB107	2	G→A	31119	R[333]→Q	Catalytic	Yes	P	
HB108	5[c]	A→T	31161	N[347]→I	Catalytic	No	N	
HB124	<1	G→A	31165	M[348]→I	Catalytic	No	G	
HB87	<1	G→A	31218	G[366]→E	Catalytic	No	G	
HB86	<1[c]	A→G	31281	E[387]→G	Catalytic	No	P	
HB92	<1	T→C	31340	W[407]→R	Catalytic	No	G	Koeberl et al. 1989
HB66	3	C→A	31356	T[412]→K	Catalytic	No	S	

[a] Mutations at the dinucleotide CpG. All but R[333]→L were transitions.

[b] See Methods for definitions. G = conserved; P = partially generic; S = factor IX specific; and N = nonconserved.

[c] Posttherapy values but clinically severe by criteria of Eyster et al. (1980).

[d] Values measured by our laboratory. Otherwise, values are as reported by the referring hemophilia centers.

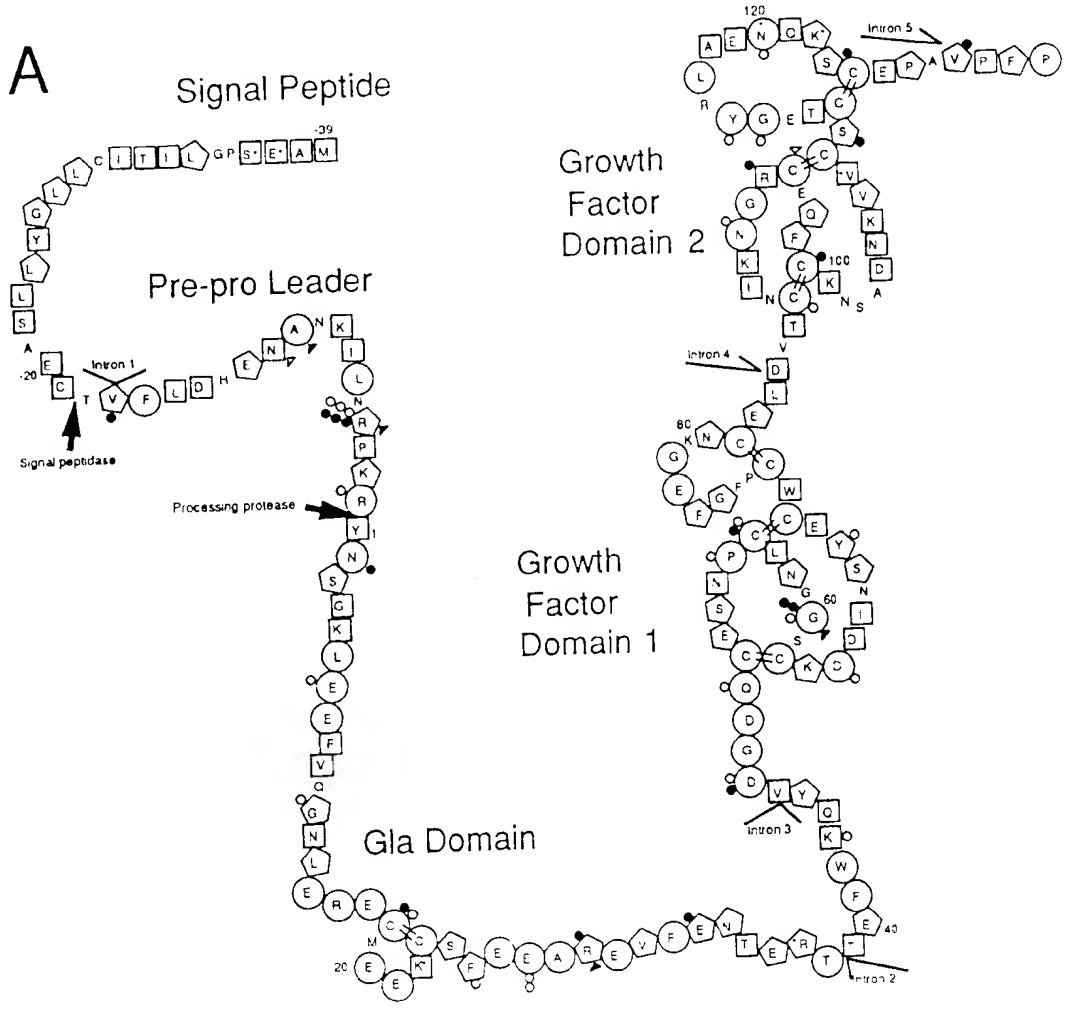
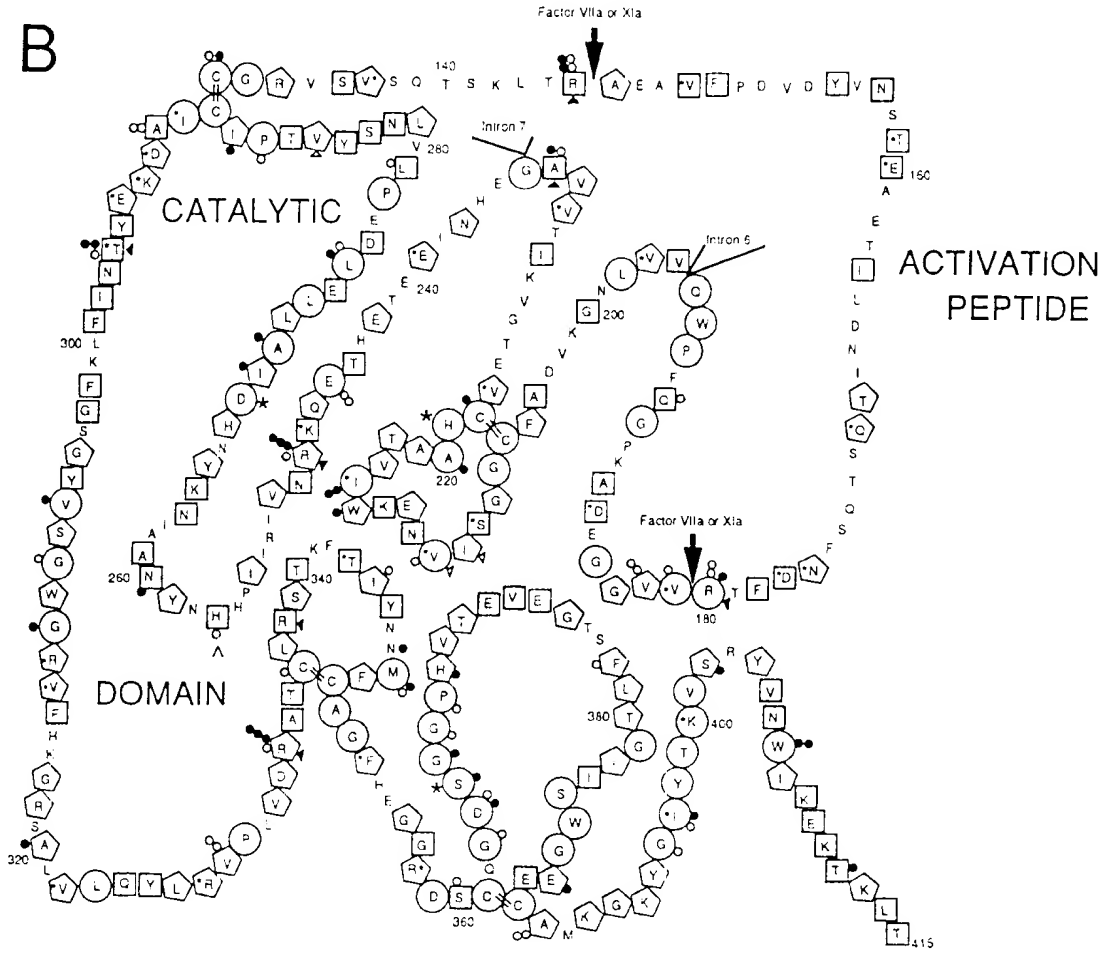


Figure 1 Factor IX missense mutations and amino acid conservation. Both the conservation of the amino acid residues in factor IX and the location of missense mutations are shown. ● = Missense mutation at a given residue from our sample; ○ = missense mutation at a given residue reported by others (Giannelli et al. 1990). Multiple symbols (● or ○) indicate the number of known independent missense-mutation changes at a given residue. For CpG dinucleotides, symbols are as follows: ▲ = conserved residues in which transitions *and* transversions will cause a nonconservative missense substitution; △ = conserved residues where *only* transversions will cause a nonconservative missense substitution. A = Missense mutation not causing disease (Montandon et al. 1990). The geometric shapes indicate the degree of conservation of an amino acid residue, as determined from factor IX and related serine protease alignments (Appendixes A and B). Circles represent "generic" residues, squares represent "factor IX-specific" residues, and pentagons represent "partially generic" residues as defined in Methods. An asterisk (*) inside the geometric shape indicates that conservative substitutions occurred in the mammalian species of factor IX. The alignment of factor IX is based on the available mammalian species of factor IX and on both the human and bovine sequences of factor X, factor VII (Takeya et al. 1988), and protein C (Koeberl et al. 1990; Sarkar et al. 1990) (see Appendixes A and B). *A*, Alignment of N-terminal segment of factor IX, based on four available mammalian species: human (Yoshitake et al. 1985), cow (Katayama et al. 1979), dog (Evans et al. 1989), and mouse (Wu et al. 1990). The translation start site is based on additional data from rat and macaque sequences (Pang et al. 1990). *B*, Alignment of C-terminal segment of factor IX (activation peptide and catalytic domain), based on nine mammalian species: human, sheep, pig, rabbit, guinea pig, rat, mouse, cow, and dog (Evans et al. 1989; Sarkar et al. 1990). The single-letter code for amino acids is as follows: A = Ala; R = Arg; N = Asn; D = Asp; C = Cys; Q = Gln; E = Glu; G = Gly; H = His; I = Ile; L = Leu; K = Lys; M = Met; F = Phe; P = Pro; S = Ser; T = Thr; W = Trp; Y = Tyr; and V = Val. Stars (★) indicate the serine protease catalytic triad amino acids. Note that the classification in table 3 is based on factor IX from only four species and that only non-CpG residues were tabulated.



ent haplotype, indicating that these two mutations are due to independent events.

Missense Mutations versus Extent of Evolutionary Conservation

To determine the relationship between missense mutations in hemophiliacs and the amino acid conservation classes, haplotype analysis was used to determine whether recurrent mutations were independent events. Mutations at non-CpG dinucleotides were considered separately from those at CpG dinucleotides, since mutations at CpG dinucleotides are greatly enhanced. Although the CpG category represents less than 3% of all the missense mutations that are possible, the analysis of this category is important (see below).

At least 38 independent missense mutations were found at non-CpG residues (table 1; Bottema et al. 1990*a*; Koeberl et al. 1990). We combined these data with data on 43 additional non-CpG mutations reported in the literature (reviewed in Giannelli et al. 1990). Since haplotype data are not generally available on patients reported by others, it was not possible to identify recurrent mutations that are due to independent mutational events. However, in our sample, extensive haplotype analysis strongly suggests that only one of the 25 recurrent mutations at *non-CpG* dinucleotides is independent (Bottema et al. 1989, 1990*b*; Ketterling et al. 1991, and in press). Thus, it seems reasonable to assume that all recurrent mutations reported in the literature are the result of common ancestors. Common ancestry has been well docu-

mented for the most frequently recurring non-CpG mutation: isoleucine[397]→threonine (Bottema et al. 1990*b*; Thompson et al. 1990).

Four classes of factor IX residues can be defined on the basis of the extent of evolutionary conservation. These classes are (1) generic residues that are conserved in factor IX and in the related coagulation serine proteases factor VII, factor X, and protein C, (2) partially generic residues that are conserved in one or two of the related coagulation proteases, (3) factor IX–specific residues that are conserved in factor IX but in none of the related coagulation proteases, and (4) nonconserved residues (see Methods).

The amino acids in factor IX are distributed between the four classes in a nonrandom manner (table 2). Almost one-half of the generic residues are either cysteine, glycine, or glutamate. The cysteine residues are involved in disulfide bonds (fig. 1), and most of these glutamate residues are modified to γ-carboxyglutamic acids necessary for calcium binding (Furie and Furie 1988) (table 2). A substantial fraction of the charged generic residues (10 glutamates and three aspartates) are involved in the chelation of calcium (Rees et al. 1988). If calcium binding is ignored, charged residues are substantially underrepresented in the generic and partially generic classes. In contrast, codons having a high G + C content are substantially overrepresented (Bottema et al. 1991). If the generic and partially generic residues are combined, cysteine, glycine, and tryptophan are significantly overrepresented while threonine, asparagine, and lysine are underrepresented (for χ^2 values, see table 2, footnotes b and c).

Both the classification of each residue in factor IX and the location of all the missense mutations observed in patients with hemophilia B were determined (fig. 1). The non-CpG missense mutations are distributed throughout the factor IX protein. However, missense mutations causing hemophilia B are most likely to occur at generic residues (table 3). Mutations at the partially generic residues are about twofold less likely to produce hemophilia B; and mutations at the factor IX–specific residues are sixfold less likely to produce hemophilia B (i.e., the likelihood that they will produce the disease is only 15% of that for generic residues (table 3). Mutations at nonconserved residues are 33-fold less likely to produce hemophilia B, indicating that these residues are almost never involved in this disease.

Three models for these observations can be envisioned. These models will be stated in the context of possible explanations for the low frequency of causative missense mutations at factor IX–specific residues:

1. At 15% of the factor IX–specific residues, *most*, if not all, of the possible amino acid substitutions will cause disease. The remaining 85% of the factor IX–specific residues are not essential. *In this case, factor IX sequence from additional nonmammalian species should allow sufficient evolutionary*

Table 2

Number of Residues in Each Class for Human Factor IX

Residue Type[a]	No. of Generics	No. of Partial Generics	No. of Factor IX Specifics	No. of Nonconserved
Nonpolar:				
A	5	3	9	7
V	4	17	7	8
L	7	9	7	5
I	3	7	9	5
P	7	1	2	5
F	3	8	6	4
W[b]	4	2	1	0
M	1	0	1	2
Subtotal	34	47	42	36
Polar				
G[b]	19	10	4	3
S	2	10	5	9
T	2	4	14	10
C[b]	22	0	1	1
Y	2	6	8	0
N[c]	3	5	12	11
Q	2	3	4	4
Subtotal	52	38	48	38
Charged:				
D	6	2	7	4
E/γ[d]	11 (9)	10 (3)	11 (0)	11 (0)
K[c]	1	7	12	8
R	3	10	3	3
H	1	2	1	6
Subtotal	22	31	34	32
Total	108	116	124	106

SOURCE.—Fig. 1.

[a] As categorized by Lehninger (1975).

[b] More abundant in generic and partially generic classes than in factor IX–specific and nonconserved classes. By the χ^2 test, $P < .001$ for cysteine and glycine and $P < .05$ for tryptophan.

[c] More abundant in factor IX–specific and nonconserved classes than in generic and partially generic classes. By the χ^2 test, $P < .001$ for threonine and $P < .03$ for asparagine and lysine.

[d] γ = γ-carboxyglutamic acid. The numbers in parentheses are the numbers of modified residues.

Table 3

Frequency of Missense Changes in Factor IX That Are Due to Mutations at Non-CpG Dinucleotides, as Function of Amino Acid Conservation

	Distribution, by Amino Acid Conservation Class, Based on Factor IX Sequences from Four Species[a]													
	Generic			Partially Generic			Factor IX Specific				Nonconserved			
	N Terminal	C Terminal	Total	N Terminal	C Terminal	Total	N Terminal	C Terminal	Total	Total Conserved	N Terminal	C Terminal	Total	Total
A. No. of residues	52	50	102	38	83	121	54	87	141	364	25	53	78	442
B. No. of missense mutations observed[b]	23	24	47	7	16	23	2	8	10	80	0	1	1	81
C. Frequency relative to generics[c] . . .			1.0			.41			.15	. . .			.03	. . .
D. Maximal target size[d] (A × C)			102			51			21	174 (48%)			2	176 (40%)[e]

Note.—The signal-through-EGF2 domains of factor IX constitute the N-terminal segment (N terminal), and the activation and catalytic domains of factor IX constitute the C-terminal segment (C terminal). Factor IX residues in four species were compared with human and bovine factor VII, human and bovine factor X, and human and bovine protein C (Appendix B). The four species of factor IX included human, mouse, cow, and dog (Appendix A).

[a] See Results for definitions of the residue classes, and see Methods for details of classification (93% of the conserved residues are identical in the four species of factor IX, and 7% are conservatively substituted).

[b] Data include our missense mutations at non-CpG nucleotides of factor IX, as well as those reported by Giannelli et al. (1990). (Five mutations [6%] occur at conservatively substituted residues.)

[c] Frequencies are normalized with respect to missense mutations/generic residues.

[d] The number of residues predicted to cause hemophilia B, if it is assumed that 100% of missense mutations at generic residues cause disease (row A × Row C).

[e] When the C-terminal classification is based on nine rather than on four mammalian factor IX sequences, the target size decreases only slightly, to 39% of total residues.

time for most nonessential residues to be substituted, while the essential 15% of functionally important residues will remain conserved.

2. Mutations at *100%* of the factor IX–specific residues can cause disease, but, on average, at any given site, only 15% of the 19 possible amino acid substitutions will cause disease, and 85% of the substitutions will not cause disease. *In this case, the number of factor IX–specific residues will approach zero as progressively more factor IX sequences are added. Thus, the additional factor IX sequences will increase the fraction of causative mutations occurring at nonconserved residues.*
3. Missense mutations at 15% of the factor IX–specific residues cause hemophilia B, and the remaining 85% cause an as yet undefined disease such as a hypercoagulability that might perhaps be lethal prenatally. *In this case, the 141 factor IX–specific residues should remain conserved if other factor IX sequences are added.*

To help distinguish between these possibilities, sequence data were analyzed from the C-terminal segment (the activation and catalytic domain) of an additional five species of factor IX (Sarkar et al. 1990). An alignment of all nine species (C-terminal 9) versus the alignment of four species (C-terminal 4) indicates that 27 residues (13%) are now reclassified as nonconserved (table 4A). None of these reclassified residues were generic in the initial C-terminal 4 alignment. In contrast to the prediction of model 3, 12% of the partially generic and 21% of the factor IX–specific residues were reclassified as nonconserved (table 4B). More important, no missense mutations have been reported to cause hemophilia B in any of the 27 reclassified nonconserved residues. If there had been a corresponding decline in the number of missense mutations at conserved residues, as predicted by model 2, an additional 3.5 mutations would be expected at nonconserved residues (table 4B). However, none were observed ($P < .002$). While the data best support the predictions of model 1, more cross-species sequences and a larger sample of mutations are necessary to eliminate a hybrid model that contains some contribution from model 2 and/or model 3.

The pattern of transitions at CpG dinucleotides further supports model 1. Since the mutation rate in factor IX is much higher at CpG dinucleotides than at non-CpG dinucleotides (Koeberl et al. 1989; Green et al. 1990), the mammalian factor IX sequences at CpG

Table 4

Amino Acid Conservation and Missense Mutations in Activation and Catalytic Domains of Factor IX at Non-CpG Residues in C-terminal 4 versus C-terminal 9

A. Amino Acid Conservation in C-terminal 4 vs. C-terminal 9

	Total Conserved		Total Nonconserved	
	C-terminal 4	C-terminal 9	C-terminal 4	C-terminal 9
No. of residues	218	191	54	81
No. of missense mutations observed	48	48	1	1

B. Conserved C-terminal 4 Amino Acids Converted to Nonconserved in C-terminal 9

	Generic	Partially Generic	Factor IX Specific	Total
No. of residues converted[a]	0	10	17	27
Residues converted in each C-terminal 4 conserved class	0%	12%	21%	12%
No. of observed missense mutations expected to become C-terminal 9 nonconserved[b]	0	4.1	2.6	6.7

NOTE.—Conservation is defined as in Methods. Conservation of the carboxy segment of factor IX was determined on the basis of an alignment of either C-terminal 4 or C-terminal 9, both alignments utilize sequence from human, mouse, cow, and dog. The C-terminal 9 alignment also includes sequence from sheep, pig, rat, guinea pig, and rabbit (Sarkar et al. 1990) (see Appendix A). Similar alignments have been published elsewhere (Sarkar et al. 1990; Wu et al. 1990).

[a] Residues conserved in the C-terminal 4 alignment that convert to nonconserved residues in the C-terminal 9 alignment.

[b] Number of observed missense mutations expected, on the basis of model 2, to become C-terminal 9 nonconserved; this number is calculated by multiplying the percent of C-terminal 4 conserved residues that convert to nonconserved in the C-terminal 9 alignment by the relative probability of missense mutation causing hemophilia B (table 3). For partially generic residues, $10 \times .41 = 4.1$, and for factor IX–specific residues $17 \times .15 = 2.6$.

dinucleotides should rapidly mutate at nonconserved amino acids. Therefore, the pattern observed in mammalian sequences at CpG dinucleotides should be *analogous to the pattern of non-CpG conserved dinucleotides that would be observed in more diverged species*. Transitions at the 15 conserved CpG nucleotides should cause a missense mutation resulting in disease (table 5). Transitions at 12 of the 15 possible CpG sites have been observed. These transitions have occurred in three of four generic residue sites, in five of six partially generic residue sites, and, most important, in four of five factor IX–specific residue sites. Thus, there is an almost perfect correlation between evolutionary conservation and missense mutations causing hemophilia B. Furthermore, there are four arginine residues (codons CGX, where X is any base except A) in which transitions at either C or G will produce a missense mutation. Mutations at all six evolutionarily conserved sites have been observed to cause hemophilia B, while neither of the two possible transitions have been observed in the one nonconserved CGX arginine residue (R^{403}).

Percent of Missense Mutations That Cause Hemophilia B

Factor IX, factor X, factor VII, and protein C diverged about 450–500 million years ago (Doolittle and Feng 1987). If a residue is identical in these proteases despite such a long period of evolutionary time, it is very likely to be absolutely essential for protein function. The essential nature of such generic residues (102 total) is supported by an analysis of the residues conserved in the C-terminal 4 of factor IX versus those in the C-terminal 9 of factor IX (table 4B). Generic residues are absent from the group of 27 amino acids that, as a result of the C-terminal 9 alignment, have been reclassified as nonconserved (table 4B). Additional support comes from an alignment of human and bovine factor VII, factor X, and protein C. One hundred six residues are identical in these proteins. If the four species of factor IX are added to the alignment, virtually all (96%) of these generic residues remain identical, despite an additional 450+ million years of evolutionary divergence. Thus, we conclude that most, if not all, possible missense mutations at generic residues will cause disease. From the relative frequencies of mutations in each class, it can be estimated that only 40% of all possible missense changes in factor IX will cause hemophilia B (table 3). In the context of model 1, the estimate implies that 40% of factor IX residues are important for function and that most, if not all, missense mutations in these residues will cause disease.

Discussion

Prediction of Missense Mutations That Cause Hemophilia

We have analyzed (*a*) amino acid changes that disrupt factor IX function in hemophiliacs and (*b*) amino

Table 5

Missense Changes in Factor IX That Are Due to Transitions at CpG Dinucleotides, as Function of Amino Acid Conservation

	Amino Acid Conservation Corresponding to C or G Nucleotide at CpG[a]					
	Generic	Partially Generic	Factor IX Specific	Total Conserved	Nonconserved	Total
A. No. of C or G nucleotides at CpG[b]	4	6	5	15	6	21
B. No. of our independent missense mutations[c]	2	7	5	14	0	14
C. No. of sites at which missense transitions have been reported[d]	3	5	4	12	0	12

[a] At certain arginine residues (CGX), a transition at either C or G will cause a missense mutation, while in other cases transitions at only G will cause a missense mutation. Thus, each transition which causes a missense mutation is counted separately.

[b] Nucleotides in which transitions result in missense mutations.

[c] Data include our independent transitions resulting in missense mutations at CpG dinucleotides of factor IX. Multiple mutations at the same site were judged as independent only if the haplotype differed or if a germ line of origin could be determined (table 1; Bottema et al. 1990*a*; Koeberl et al. 1990).

[d] Data are from table 1 and from Bottema et al. (1990*a*), Giannelli et al. (1990), and Koeberl et al. (1990).

acid changes that are compatible with normal factor IX function in different species. Four classes of factor IX amino acids were defined on the basis of the extent of evolutionary conservation. We document the functional importance of most, if not all, of the generically conserved residues. However, residues uniquely conserved in factor IX are sixfold less likely to give rise to hemophilia B. Thus, many mutations at factor IX–specific residues should be neutral variants. One such neutral variant in a factor IX–specific residue has recently been discovered: histidine^{257} is nonconservatively changed to tyrosine without causing hemophilia B in a male (Montandon et al. 1990).

Both (1) the relationship between missense mutation and amino acid conservation class as defined by the carboxy segment of factor IX in C-terminal 4 versus that in C-terminal 9 and (2) an analysis of missense mutations at CpG dinucleotides suggest that about 40% of the residues in factor IX are crucial for function. Most of the possible missense mutations in the remaining 60% of the residues will not cause hemophilia B; these remaining residues are likely to be "spacers," i.e., residues which maintain the position of critical amino acids but whose own side chains are not crucial for function (Doolittle and Blombäck 1964). Many of these spacer residues are classified as factor IX specific because the evolutionary time separating the mammalian factor IX sequences is insufficient to have changed many nonessential residues. The conclusions predict that nonmammalian factor IX sequences should convert many of these nonessential residues to the nonconserved class and substantially increase the likelihood that a mutation in the remaining factor IX–specific residues will cause disease.

Saturation in vitro mutagenesis of selected residues in factor IX and expression in tissue culture could help confirm that factor IX function will be significantly compromised by the amino acid substitutions found in hemophiliacs. However, the generation, confirmation, and characterization of such a large number of mutations would require many years of effort. In addition, the interpretation of data indicating that a mutation retains functional integrity is confounded by the simplicity of a cell culture system in comparison with the intact organism (i.e., mutants that score as functional in cell culture could still cause hemophilia B in humans). It would be preferable to generate a battery of transgenic mice with hemophilia B, but both the absence of a mouse model for hemophilia B and the difficulties in generating a large number of transgenic mice with independent mutations pose major technical challenges. We conclude that sequencing factor IX in more nonmammalian species and delineating the mutations in a larger sample of hemophiliacs is currently the best way to determine which missense mutations will result in disease.

The generality of the present findings can be assessed by examining the correlation between evolutionary conservation and mutations causing other severe X-linked diseases. Hemophilia A would be a good choice, because factor VIII belongs to a different gene family and more than 100 missense mutations will soon be available. Autosomal genes such as α- and β-globin are *not* good candidates for assessing the relationship between evolutionary conservation and missense mutations that disrupt function. In the cases of α- and β-globin, the marked overrepresentation of dominant mutations, heterozygote advantage, founder effect, and the biased methods of patient ascertainment pose major problems in the interpretation of the data. As an example of the problems, the aggregate mutational data in globin erroneously indicated that CpG was *not* a hot spot of mutation (Vogel and Motulsky 1986).

Mutant Analyses in Other Genes

The data, albeit meager, from saturation in vitro mutagenesis in other systems is compatible with the notion that, if one missense mutation at a residue disrupts function, then the other possible missense mutations are also very likely to disrupt function. In *Escherichia coli,* saturation mutagenesis of evolutionarily conserved residues in the region of the β-lactamase active site revealed that 14 of the 19 possible amino acid substitutions retained appreciable activity toward the penicillins (Schultz and Richards 1986). However, in all but two of the substitutions the limited characterization performed was sufficient to reveal major reductions in catalytic specificity and/or thermal stability, strongly suggesting that all these mutants would be at a selective disadvantage in vivo. Moreover, in a follow-up study using saturation mutagenesis at five codons, partial activity could commonly be found, but no mutant protein had the catalytic specificity and thermal stability of a wild-type protein (Dube and Loeb 1989). Finally, in NIH 3T3 cells, a study of substitutions at the conserved glycine^{12} of the Harvey ras protein indicated that 18 of 19 amino acid substitutions produced a transformed phenotype (Seeburg et al. 1984).

Analysis of the N-terminal segment of the lambda repressor by cassette mutagenesis indicates both some

residues in which only the wild-type side chain is acceptable and other residues in which either a few or many substitutions are acceptable (Reidhaar-Olson and Sauer 1988; Lim and Sauer 1989). However, interpretation of the data is complicated by (1) the generation of multiple potentially compensatory mutations by cassette mutagenesis, (2) the biases associated with a mutation method which relies on equal pairing of inosine with all bases, (3) the use of only the N-terminal fragment of the repressor, and (4) the use of a selection or screening scheme without knowledge of how that translates into the fitness of the viral protein in its ecosystem.

Caveats

Multiple base substitutions at a single codon are rare, and no such missense mutations have yet been reported in the factor IX gene. If the present pattern is representative of the past, a residue will be conserved through evolution if the five to eight possible single-base missense changes all cause disease. It is conceivable that disease will not be caused by missense changes that arise from multiple base substitutions at a codon. However, this seems unlikely because 94% of the generic residue sites do not tolerate even highly conservative substitutions which usually involve a single-base change (see Methods and fig. 1). Therefore, the more drastic missense changes that commonly result from substitutions at two and three bases are unlikely to be tolerated.

A second caveat concerns the virtual certainty that at least a small fraction of residues may fit model 2 and perhaps model 3. Such residues will limit the extent to which evolutionary conservation can predict which mutations will cause hemophilia B in humans. Both identification of more missense mutations and additional sequencing of factor IX from nonmammalian species should ultimately allow an estimate of the fraction of residues fitting models 2 and 3.

Possible Implications for Clinical Research

The development of rapid PCR-based methods for direct sequencing and screening assures that many protein sequence variants will be detected by the analysis of DNA. Some of these variants will be found in individuals who also carry a normal allele. How does one assess the likelihood of the change being neutral, as opposed to a change that either predisposes to a multifactorial disease in heterozygotes or causes a recessively inherited disease in homozygotes? Given the expense and effort of clinical studies, it would be useful to have criteria for estimating the likelihood that a missense mutation observed in a heterozygote will produce a dysfunctional protein. If further data were to show that the present observations are generally true, the level of evolutionary conservation might provide such criteria.

Acknowledgments

We thank Joslyn Cassady, Renee McGovern, and Amy Groszbach for contributing to the sequencing; Charyl Dutton, Christine Paynton, and Gobinda Sarkar, Ph.D., for contributing to the haplotype analysis; Darrell Ricke and Paul Axelsen, M.D., for help with the conservation data alignment; and Mary Johnson for excellent secretarial assistance. We thank Karin Blakemore, M.D., and Virginia Corson, M.S., both of John Hopkins Hospital; John Rinehart, M.D., and Pat Burkey, R.N., both of Ohio State University Hospital; Joan Gill, M.D., and Luann Weik, M.S., both of Children's Hospital of Wisconsin; Roger Edson, M.D., and Bonnie LeRoy, M.S., both of University of Minnesota Hospital; Charles Sexauer, M.D., and Joyce Shissler, R.N., both of Oklahoma Children's Memorial Hospital; Marian Koerper, M.D., and Pat Liddy, R.N., both of University of California, San Francisco Hemophilia Center; Carol Kaspar, M.D., and Peter Mockary, M.S., both of Orthopedic Hospital Hemophilia Center of Los Angeles; Bruce Furie, M.D., and Steven Limentani, M.D., both of New England Medical Center; and Emily Czapek, M.D., and DeAnne Schmid, R.N., both of Regional Comprehensive Hemophilia Center of Central and Northern Illinois, for providing pedigrees and blood samples. We thank Kim Maness and Melinda Cohen of Vanderbilt University for providing DNA samples, laboratory values, and pedigrees. This work was aided by grant 5-647 from the March of Dimes—Birth Defects Foundation and by NIH grant HL39762. We thank Paul Axelsen, M.D., Whyte Owen, Ph.D., and David Fass, Ph.D., for helpful discussions.

Appendix A

Alignment of N-Terminal Segment of Factor IX

```
                                        *Intron 1
                    -30v       -20v       -10v        +1v
Human       MAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSG
Cow                                                 ....
Mouse       ....RA....F.......T..A....R...T...T........
Dog         ...AS..............A....R...T...S........
Rat         ..DAP..
Macaque     .......

                                                *Intron 2*Intron 3
                10v        20v        30v        40v        50v
Human       KLEEFVQGNLERECMEEKCSFEEAREVFENIERTIEFWKQYVDGDQCESN
Cow         ......R.......K.................K.................
Mouse       ......R.......I..R..............K.................
Dog         ......R.......I..R..............K.................

                                                    *Intron 4
                60v        70v        80v        90v        100v
Human       PCLNGGSCKDDINSYECWCPFGFEGKNCELDVICNIKNGRCEQFCKNSAD
Cow         ......M............QA....T.....A..S......K....RDT.
Mouse       ......I.....S......QV....R.....A.........K......P.
Dog         ....D.V............RA....................K....LGP.

                                        *Intron 5
                110v       120v       130v
Human       NKVVCSCTEGYRLAENQKSCEPAVPFPCG
Cow         ........D......D.............
Mouse       ...I........Q...D......T......
Dog         ........T..Q...D.R...........
```

Figure A1 Amino acid sequences of factor IX from mammalian species, aligned from published sequences. The full-length proteins were aligned from human (Yoshitake et al. 1985), cow (Katayama et al. 1979), dog (Evans et al. 1989), and mouse (Wu et al. 1990). For the activation and catalytic domains, additional sequence from sheep, guinea pig, rat, pig, and rabbit (Sarkar et al. 1990) was used for the analyses that compared nine with four species (table 4). The translation start site is based on additional data from rat and macaque (Pang et al. 1990). The factor IX–specific residues are underlined.

```
                 140v       * 150v       160v                  170v
Human            RVSVSQTSK-LTR AEAVFPDVDYVNSTEA----------ETILDNITQST
Sheep            .A..LH...K... ..TI.SNMN.E..S..----------.I.W..V...N
Pig                ..HSPTT... ..II.SNM..E....V----------.P...SL.E.N
Rabbit           .....HA..KI.. .TTI.SNTE.E.F...----------...RG.V..RS
Guinea Pig       ...IPSV..EHN. .N.I.SRMG...F.DDETIWDDNDDD...W..S.E..
Rat              ....AYN..KI.. ..T..SNT..G....L--ILDDITN-S.....L.ENS
Mouse            .A.I.YS..KI.. ..T..SNM..E.....VFIQDDITD-GA..N.V.E.S
Cow              .....HI..K... ..TI.SNTN.E..S..----------.I.W..V...N
Dog              ....PHI.MTR.. ..TL.SNM..E....V----------.K....V..--

                                        *Intron 6
                  180v*      190v       200v       210v       220v
Human            QSFNDFTR VVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHC
Sheep            ...D..N. .......AR.......L.H.EIA............V......
Pig              ..SD..I. I....N.........L....I........I....V......
Rabbit           ..SD.... I....N.........L.....E.......I....V......
Guinea Pig       KPSDE.F. ................L...ETE...................
Rat              EPI..... .....N.....I....I...EIE.....A.I...........
Mouse            E.L..... .....N.....I....I...EIE.....A.I...........
Cow              ...DE.S. .......ER.......L.H.EIA............V......
Dog              -PL..... ....K...........L.............I....V......

                                   *Intron 7
                  230v       240v       250v       260v       270v
Human            VETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIAL
Sheep            IKP............T.KP.P.........A..Y.G...S....S.....
Pig              I.P..........Y.T....P...R.....A....S...TV...S.....
Rabbit           IKPDDN.......Y..Q...N............Y.K...T..........
Guinea Pig       ILP.I..E....K....KK.D...R...TQ..L..S...SF...S.....
Rat              LKP.D..E........DEK.D...R.....T....Q...T....S.....
Mouse            LKP.D..E.....Y..DKK.D...R.....T....Q...T....S.....
Cow              IKP............T.KP.P.........A..Y.S...S....S.....
Dog              I.PD....I......T.KR...........T.L..S...T..........

                  280v       290v       300v       310v       320v
Human            LELDEPLVLNSYVIPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALV
Sheep            .......E............R..........Y..........NR....SI
Pig              .......T.....................................NR....TI
Rabbit           ....K..T............NR.......N.............NR..Q.SI
Guinea Pig       ....K..S............NR..........A.......KL.SQ..T.SI
Rat              ....K..I.........V.N.....................K..N...Q.SI
Mouse            ....K..I.........V.NR....................K..N...Q.SI
Cow              .......E............RD.....S...Y.......K..NR....SI
Dog              .......T............R..S.....................N.....SI
```

Appendix A (continued)

```
                330v        340v        350v        360v        370v
Human           LQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTE
Sheep           ....K..................H.....Y....K..............
Pig             ....K.............V...S..........K...L..........
Rabbit          .......F.......................DV..K...E..........
Guinea Pig      ..................................................
Rat             ..................S.........YR...K...E..........
Mouse           ...................T..........YR...K...E..........
Cow             ....K..............S..SH.....Y....K..............
Dog             ....K................................K..............
```

```
                380v        390v        400v        410v
Human           VEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT
Sheep           ..............................
Pig             ..................V...............
Rabbit          ..................I.....V..R..W....
Guinea Pig      ....N.............................
Rat             ..................................
Mouse           ..........................................
Cow             ..........................................
Dog             ...I......................................
```

Appendix B

Alignment of Related Coagulation Serine Proteases

```
                          -39                                  * Intron 1
Human Factor IX              MAE SPGLITICLL GYLLSAECTV FLDHENANKI LNRPKRYNSG
Bovine Factor IX                                                        ....
Human Factor X            MGRPL HLV.LSAS.A .L..LG.-SL .IRR.Q..N. .A.VT.A..-
Bovine Factor X           MAGLL HLV.LSTA.G .L.RPAG-S. ..PRDQ.HRV .Q.AR.A..-
Human Factor VII             MV .QA.RLL... LG.QGCLAA. .VTQ.E.HGV .H.RR.A.A-
Bovine Factor VII                                                     A.-.
Human Protein C           MWQLTS LLLFVATWGI SGTPAPLDS. .SSS.R.HQV .RIR..A..-
Bovine Protein C             TS LLLFVT.WGI SSTPAPPDS. .SSS.R.HQV .RIR..A..-

                5                                          *Intron 2*Intron 3
H. Factor IX  KLEEFVQGNL EREGMEEKCS FEEAREVFEN TERT TEFWKQ YVDGDQC----ESN
B. Factor IX  ......R... ....K..... .......... ..K. ...... .......----...
H. Factor X   F...MKK.H. .......T.. Y........D SDK. N...NK .K.....----.TS
B. Factor X   F...VK.... ....L..A.. L........D A.Q. D...SK .K.....----.GH
H. Factor VII F...LRP.S. ....K..Q.. ......I.KD A... KL..IS .S.....----A.S
B. Factor VII F...LRP.S. ....R..L.. ....H.I.R. E... RQ..VS .N.....----A.S
H. Protein C  F...LRHSS. ....I..I.D ....K.I.Q. VDD. LA..SK H......LVLPLEH
B. Protein C  F...LRP..V ....S..V.E ......I.Q. ..D. MA..SK .S.....EDRPSGS

               55                                          * Intron 4
H. Factor IX  PCL----NGGSCKD DINSYECWCP FGFEGKNCEL DVT----CNIKNGR CEQFCKNSAD
B. Factor IX  ..------...M... .........Q A....T.... .A.----.S..... .K....RDT.
H. Factor X   ..Q----.Q.K... GLGE.T.T.L E......... FTRK--L.SLD..D .D...-HEEQ
B. Factor X   ...----.Q.H... G.GD.T.T.A E........F STRE--I.SLD..G .D...-REER
H. Factor VII ..Q---.....    QLQ..I.F.L PA...R...T HKDDQLI.VNE..G ...Y.SDHTG
B. Factor VII ..Q---.....E.  QLR..I.F.. D....R...T .KQSQLI.AND..G ...Y.GADPG
H. Protein C  ..ASLCCGH.T.I. G.G.FS.D.R S.W..RF.QR E.SF-LN.SLD..G .THY.-LEEV
B. Protein C  ..DLPCCGR.K.I. GLGGFR.D.A E.W..RF.LH E.RF-SN.SAE..G .AHY.-MEEE

              105                          * Intron 5
H. Factor IX    NKVVCSCTEG YRLAENQKSC EPAVPFPCGR VSVSQTSK-LT RAEAVFPDVD YVNSTEAETI
B. Factor IX    ........D. .....D.... .......... ....HI..K.. ...TI.SNTN .E..S...I.
H. Factor X     .S.....AR. .T..D.G.A. I.TG.Y...K QTLERRKRSVA Q.TSSSGEAP DSITWKPYDA
B. Factor X     SE.R...AH. .V.GDDS.S. VSTER....K FTQGR---SSR W.IHTSEDAL DASEL.HYDP
H. Factor VII   T.RS.R.H.. .S.LADGV.. T.T.EY...K IPILEKR---- ---------- ----------
B. Factor VII   AGRF.W.H.. .A.QADGV.. A.T.EY...K IP.LEKR---- ---------- ----------
H. Protein C    GWRR...AP. .K.GDDLLQ. H...K..... PWKRMEK.RSH LKRDTE---- ----------
B. Protein C    GRRH...AP. ...EDDHQL. VSK.T..... LGKRMEK.RK. LKRDTN---- ----------
```

Figure B1 Factor IX sequences aligned with human and bovine sequences from related coagulation serine proteases factor VII (Hagen et al. 1986; Takeya et al. 1988), factor X (Fung et al. 1984, 1985), and protein C (Long et al. 1984; Beckmann et al. 1985). The generic and partially generic residues are underlined.

Appendix B (continued)

```
                                                                                *Intron 6
            165
H. Factor IX    LD----------NITQSTQS FN--DFTRVVGG EDAKPGQFPW Q VVL-NGKVDA
B. Factor IX    W.----------.V...N.. .D--E.S..... ...ER..... . .L.-H.EIA.
H. Factor X     DLDPTENPFDLLDF.Q..PERG D.--NL..I... QEC.D.EC.. . AL.I.EENEG
B. Factor X     DLSPTESSLDLLGL.R.EPSAG EDGSQVV.I... R.CAE.EC.. . AL.V.EENEG
H. Factor VII   ----------NA SK--PQG.I... KVCPK.EC.. . .L.-LVNGAQ
B. Factor VII   ----------NG SK--PQG.I... HVCPK.EC.. . AM.-KLNGAL
H. Protein C    ----------D.E DQ--VVP.LID. KMTRR.DS.. . ...LDS.KKL
B. Protein C    ----------.VD.K DQ--LVP.I.D. QE.GW.ES.. . A..LDS.KKL


            205                                          *Intron 7
H. Factor IX    FCGGSIVNEK WIVTAAHC---VE TGVKITVVAG EHNIEETEHT EQKRNVIRII
B. Factor IX    .......... .V.........IK P......... ...T.KP.P. ........A.
H. Factor X     ....T.LS.F Y.L.........LY QAKRFK.RV. DR.T.QE.GG .AVHE.EVV.
B. Factor X     ....T.L..F YVL.........LH QAKRF..RV. DR.T.QE.GN .MAHE.EMTV
H. Factor VII   L...TLI.TI .V.S....FDKIK NWRNLIA.L. ..DLS.HDGD ..S.R.AQV.
B. Factor VII   L...TL.GPA .V.S....FERLR SRGNL.A.L. ..DLSRV.GP ..E.R.AQ..
H. Protein C    A..AVLIHPS .VL.....---MD ESK.LL.RL. .YDLRRW.KW .LDLDIKEVF
B. Protein C    V..AVLIHVS .VL.V...---LD SRK.LI.RL. .YDMRRW.SW .VDLDIKEV.


            255
H. Factor IX    PHHNYNAAIN KYNHDIALLE LDEPLVLNSY VTPICIADKE YTNIFL--KFGS
B. Factor IX    .Y.S...S.. ..S....... .....E.... ........RD .....S-...Y
H. Factor X     K.NRF--TKE T.DF...V.R .KT.ITFRMN .A.A.LPERD WAEST.-MTQKT
B. Factor X     K.SRF--VKE T.DF...V.R .KT.IRFRRN .A.A.LPE.D WAEAT.-MTQKT
H. Factor VII   IPST.--VPG TT.......R .HQ.V..TDH .V.L.LPERT FSERT.-AFVRF
B. Factor VII   VPKQ.--VPG QTD..V...Q .AQ.VA.GDH .A.L.LP.PD FADQT.-AFVRF
H. Protein C    V.P..--SKS TTDN.....H .AQ.AT.SQT IV...LP.SG LAERE.NQAGQE
B. Protein C    I.P..--TKS TSDN.....R .AL.AT.SQT IV...LP.SG LSERK.TQVGQE


            305
H. Factor IX    GYVSGWG----RVF HKGRS-ALVLQ YLRVPLVDRA TCL-----RSTKFTI YNNMFCAGFH
B. Factor IX    ..........----K.. NR....-.SI.. ..K..... ..........S. .SH.....Y.
H. Factor X     .I...F.-----.TH E...Q-STR.K M.E..Y...N S.K-----L.SS.I. TQ......YD
B. Factor X     .I...F.-----.TH E...L-SST.K M.E..Y...S ..K-----L.SS... TP......YD
H. Factor VII   SL.........-QLL DR.AT-..E.M V.N..RLMTQ D..QQSRKVGDSPN. TEY.....YS
B. Factor VII   SA.........-QLL ER.VT-.RK.M VVL..RLLTQ D..QQSRQ.PGGPVV TD......YS
H. Protein C    TL.T...YHSS.EK EAK.NRTF..N FIKI.V.PHN E.S-----EVMSNMV SE..L...IL
B. Protein C    TV.T...----YRD ETK.NRTF..S FIK..V.PYN A.V-----HAMENK. SE..L...IL


            355
H. Factor IX    EGGRDSCQGD SGGPHVTEVE GTSFLTGIIS WGEECAMKGK YGIYTKVSRY VNWIKEKTKL T
B. Factor IX    ...K...... .......... .......... .......... .......... .......... .
H. Factor X     TKQE.A.... .......RFK D.Y.V...V. ...G..R... .......TAF LK..DRSM.T R
B. Factor X     TQPE.A.... .......RFK D.Y.V...V. ...G..R... F.V.....NF LK..DLIM.A R
H. Factor VII   D.SK...K.. .....A.HYR ..WY....V. ..QG..TV.H F.V..R..Q. IE.LQKLMRS E
B. Factor VII   D.SK...K.. .....A.RFR ..W....VV. ...G..AA.H F....R.... TA.LRQLMGH P
H. Protein C    GDRQ.A.E.. ....M.ASFH ..W..V.LV. ...G.GLLHN ..V....... LD..HGHIRD K
B. Protein C    GDP..A.E.. ....M..FFR ..W..V.LV. ...G.GRLYN ..V....... LD..YGHI.A Q
```

References

Beckmann RJ, Schmidt RJ, Santerre RF, Plutzy J, Crabtree GR, Long GL (1985) The structure and evolution of a 461 amino acid human protein C precursor and its messenger RNA, based upon the DNA sequence of cloned human liver cDNAs. Nucleic Acids Res 13:5233–5247

Bottema CDK, Bottema MJ, Ketterling RP, Yoon H-S, Janco RL, Phillips III JA, Sommer SS (1991) Why does the human factor IX gene have a G + C content of 40%? Am J Hum Genet 49:839–850

Bottema CDK, Ketterling RP, Yoon H-S, Sommer, SS (1990*a*) The pattern of factor IX germ-line mutation in Asians is similar to that of Caucasians. Am J Hum Genet 47:835–841

Bottema CDK, Koeberl DD, Ketterling RP, Bowie EJW, Taylor SAM, Lillicrap D, Shapiro A, et al (1990*b*) A past mutation at isoleucine^{397} is now a common cause of moderate/mild haemophilia B. Br J Haematol 75:212–216

Bottema CDK, Koeberl DD, Sommer SS (1989) Direct carrier testing in 14 families with hemophilia B. Lancet 2: 526–529

Camerino G, Grzeschik KH, Jaye M, DeLaSalle H, Tolstoshev P, Lecocq JP, Heilig R, et al (1984) Regional localization on the human X chromosome and polymorphism of the coagulation factor IX gene (hemophilia B locus). Proc Natl Acad Sci USA 81:498–502

Doolittle RF, Blomback B (1964) Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. Nature 202:147–152

Doolittle RF, Feng DF (1987) Reconstructing the evolution of vertebrate blood coagulation from a consideration of the amino acid sequences of clotting proteins. Cold Spring Harbor Symp Quant Biol 52:869–874

Dube DK, Loeb LA (1989) Mutants generated by the insertion of random oligonucleotides into the active site of the β-lactamase gene. Biochemistry 28:5703–5707

Evans JP, Watzke HH, Ware JL, Stafford DW, High KA (1989) Molecular cloning of a cDNA encoding canine factor IX. Blood 74:207–212

Eyster ME, Lewis JH, Shapiro SS, Gill F, Kajani M, Prager D, Djerassi I, et al (1980) The Pennsylvania hemophilia program 1973–1978. Am J Hematol 9:277–286

Fung MR, Campbell RM, MacGillivray RTA (1984) Blood coagulation factor X mRNA encodes a single polypeptide chain containing a prepro leader sequence. Nucleic Acids Res 12:4481–4492

Fung MR, Hay CW, MacGillivray RTA (1985) Characterization of an almost full-length cDNA coding for human blood coagulation factor X. Proc Natl Acad Sci USA 82: 3591–3595

Furie B, Furie BC (1988) The molecular basis of blood coagulation. Cell 53:505–518

Giannelli F, Green PM, High KA, Lozier JN, Lillicrap DP, Ludwig M, Olek K, et al (1990) Haemophilia B: database of point mutations and short additions and deletions. Nucleic Acids Res 18:4053–4059

Green PM, Montandon AJ, Bentley DR, Ljung R, Nilsson IM, Giannelli F (1990) The incidence and distribution of CpG→TpG transitions in the coagulation factor IX gene: a fresh look at CpG mutational hotspots. Nucleic Acids Res 18:3227–3231

Gustafson S, Proper JA, Bowie EJW, Sommer SS (1987) Parameters affecting the yield of DNA from human blood. Anal Biochem 165:294–299

Hagen FS, Gray CL, O'Hara P, Grant FJ, Saari GC, Woodbury RG, Hart CE, et al (1986) Characterization of a cDNA coding for human factor VII. Proc Natl Acad Sci USA 83:2412–2416

Katayama K, Ericsson LH, Enfield DL, Walsh KA, Neurath H, Davie EW, Titani K (1979) Comparison of amino acid sequence of bovine coagulation factor IX (Christmas factor) with that of other vitamin K–dependent plasma proteins. Proc Natl Acad Sci USA 76:4990–4994

Ketterling RP, Bottema CDK, Koeberl DD, Ii S, Sommer SS. T^{296}→M, a common mutation causing mild hemophilia B in the Amish and others: founder effect, variability in factor IX activity assays and rapid carrier detection. Hum Genet (in press)

Ketterling RP, Bottema CDK, Phillips JP III, Sommer SS (1991) Evidence that descendants of three founders comprise about 25% of hemophilia B in the United States. Genomics 10:1093–1096

Koeberl DD, Bottema CDK, Buerstedde J-M, Sommer SS (1989) Functionally important regions of the factor IX gene have a low rate of polymorphism and a high rate of mutation in the dinucleotide CpG. Am J Hum Genet 45: 448–457

Koeberl DD, Bottema CDK, Ketterling RP, Bridge PJ, Lillicrap DP, Sommer SS (1990) Mutations causing hemophilia B: direct estimate of the underlying rates of spontaneous germ-line transitions, transversions, and deletions in a human gene. Am J Hum Genet 47:202–217

Lehninger AL (1975) Biochemistry: the molecular basis of cell structure and function, 2d ed. Worth, New York

Lim WA, Sauer RT (1989) Alternative packing arrangements in the hydrophobic core of lambda repressor. Nature 339:31–36

Long GL, Belagaje RM, MacGillivray RTA (1984) Cloning and sequencing of liver cDNA coding for bovine protein C. Proc Natl Acad Sci USA 81:5653–5656

Montandon AJ, Green PM, Bentley DR, Ljung R, Nilsson IM, Giannelli F (1990) Two factor IX mutations in the family of an isolated haemophilia B patient: direct carrier diagnosis by amplification mismatch detection (AMD). Hum Genet 85:200–204

Pang C-P, Crossley M, Kent G, Brownlee GG (1990) Comparative sequence analysis of mammalian factor IX promoters. Nucleic Acids Res 18:6731–6732

Rees DJG, Jones IM, Handford PA, Walter SJ, Esnouf MP,

Smith KJ, Brownlee GG (1988) The role of β-hydoxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX. EMBO J 7:2053–2061

Reidhaar-Olson JF, Sauer RT (1988) Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241:53–57

Sarkar G, Koeberl DD, Sommer SS (1990) Direct sequencing of the activation peptide and the catalytic domain of the factor IX gene in six species. Genomics 6:133–143

Schultz SC, Richards JH (1986) Site-saturation studies of β-lactamase: production and characterization of mutant β-lactamases with all possible amino acid substitutions at residue 71. Proc Natl Acad Sci USA 83:1588–1592

Seeburg PH, Colby WW, Capon DJ, Goeddel DV, Levinson AD (1984) Biological properties of human c-Ha-*ras*1 genes mutated at codon 12. Nature 312:71–75

Sommer SS, Sarkar G, Koeberl DD, Bottema CDK, Buerstedde J-M, Schowalter DB, Cassady JD (1990) Direct sequencing with the aid of phage promoters. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, New York, pp 197–205

Stoflet ES, Koeberl DD, Sarkar G, Sommer SS (1988) Genomic amplification with transcript sequencing. Science 239:414–419

Takeya H, Kawabata S-I, Nakagawa K, Yamamichi Y, Miyata T, Iwanaga S, Takao T, et al (1988) Bovine factor VII: its purification and complete amino acid sequence. J Biol Chem 263:14868–14877

Thompson AR, Bajaj SP, Chen SH, MacGillivray RTA (1990) "Founder" effect in different families with haemophilia B mutation. Lancet 1:418

Vogel F, Motulsky AG (1986) Problems and approaches. In: Human genetics. Springer, Berlin, pp 368–371

Winship PR, Anson DS, Rizza CR, Brownlee GG (1984) Carrier detection in haemophilia B using two further intragenic restriction fragment length polymorphisms. Nucleic Acids Res 12:8861–8872

Winship PR, Rees DJG, Alkan M (1989) Detection of polymorphisms at cytosine phosphoguanadine dinucleotides and diagnosis of haemophilia B carriers. Lancet 1:631–634

Wu S-M, Stafford DW, Ware U (1990) Deduced amino acid sequence of mouse blood-coagulation factor IX. Gene 86: 275–278

Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K (1985) Nucleotide sequence of the gene for human factor IX (antihaemophilic factor B). Biochemistry 24:3736–3750

Nucleic Acids Research, Vol. 20, Supplement 2027–2063

Exhibit D

Haemophilia B: database of point mutations and short additions and deletions—third edition, 1992

F.Giannelli, P.M.Green, K.A.High[1], S.Sommer[2], D.P.Lillicrap[3], M.Ludwig[4], K.Olek[4], P.H.Reitsma[5], M.Goossens[6], A.Yoshioka[7] and G.G.Brownlee[8]

Paediatric Research Unit, Guy's Tower, London Bridge, London SE1 9RT, UK, [1]School of Medicine, Division of Hematology, University of North Carolina, Chapel Hill, NC 27599, [2]Mayo Clinic, 200 Southwest First Street, Rochester, MN 55905, USA, [3]Department of Pathology, Richardson Laboratory, Queen's University, Kingston K7L 3N6, Canada, [4]Institute of Experimental Haematology and Blood Transfusion, Sigmund-Freud-Strasse 25, 5300 Bonn 1 and Institute of Clinical Biochemistry, University of Bonn, Germany, [5]Hemostasis and Thrombosis Research Unit, University Hospital, PO Box 9600, 2300 RC Leiden, The Netherlands, [6]National Institute of Health and Medical Research, 51 Av.du Maréchal de Lattre de Tassigny, 94010 Creteil, France, [7]Department of Pediatrics, Nara Medical College, Kashihara City, Nara, Japan and [8]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

The data base below lists known point mutations and short deletions and additions in the factor IX gene, causing the bleeding disorder haemophilia B or Christmas disease (for reviews, see Brownlee 1988, Giannelli 1989, Thompson 1990, Green *et al* 1991a) These mutations result in a defective clotting factor IX—a glycoprotein of 415 amino acid residues normally present in plasma and an essential component of the clotting cascade. The disease is an X-linked inherited recessive disorder affecting 1 in about 30,000 males and only very rarely females.

The purpose of this database is to update last year's one (Giannelli et al, 1991) by collecting in an accessible, summary form, molecular data on the causative mutations of haemophilia B patients worldwide. It is not intended to replace primary publications although it does contain a significant amount of unpublished work. As in previous years, we have included repeat observations of the same mutation, as well as molecularly unique mutations. We have continued our database numbering system (Giannelli *et al* (1991) giving all patients a unique Patient Identity Number (PIN or ID number).

The factor IX gene lies on the long arm of the ×chromosome at Xq27 and its entire sequence of 33 kb is known (Yoshitake et al, 1985). It contains 8 exons (a–h) encoding 6 major domains of factor IX. These are: (1) exon a—a hydrophobic *signal* peptide which targets the protein for secretion from the hepatocyte into the blood stream. (2) exons b and c—a propeptide and *gla* domain,—the latter containing 12 γ-carboxyglutamyl residues. This post-translational modification is required for the correct folding and calcium binding of factor IX. (3) exon d—a *type B, or first epidermal growth factor-like domain*, which shows homology to epidermal growth factor (EGF) and, in addition, contains conserved carboxylate residues including a β-hydroxyaspartate at amino acid 64. This domain binds an additional Ca^{2+} with high affinity (Handford et al, 1991). (4) exon e—a *type A, or second epidermal growth factor-like (EGF) domain* which lacks the conserved carboxylate residues of the EGF type B domain. (5) exon f—an *activation* domain, within which factor XIa cleaves twice, converting factor IX to IXa; (6) exons g and h—the *serine protease or catalytic domain*, responsible for the proteolysis of factor×to Xa. This region is homologous to other well studied serine proteases (e.g. chymotrypsin) and it is thought likely that his (221), asp (269) and ser (365), all participate in the classical catalytic mechanism.

Factor IX is initially synthesised in the liver as a precursor molecule, either 46, 41 or 39 amino acids (it is not known which, although 39 is probable (Pang et al, 1990)) longer at its N-terminus than the 415-long mature factor IX found in plasma. Processing steps occur in the hepatocyte prior to secretion and sequentially remove the hydrophobic signal peptide and the propeptide. In addition to the γ-carboxylàtion of the 12 N-terminal glutamyl residues carried out by a vitamin K-dependent carboxylase, and the partial β-hydroxylation of aspartate 64, N-linked carbohydrate side chains are added at residues 157 and 167 and at least an O-linked carbohydrate at serine 53.

There are 574 patient entries in this third edition of the database compared with 388 patients last year (Giannelli et al, 1991). Besides point mutants, it includes 50 short (defined as less than 20 nucleotides) deletions or additions or both, made up from 38 deletions, 9 additions and 3 examples involving both additions and deletions. There are also 12 double mutations, 1 triple mutation, 10 inhibitor patients and 3 female haemophiliacs. The list excludes 29 patients with partial or complete gene deletions or more complex rearrangements (Thompson, 1990). Of the 574 patients studied (see Summary Table), 278 are unique molecular events, the remainder being repeats. As is well known, many of these repeats occur at CG doublets and involve a CG→TG or CA change. As discussed before (Giannelli et al, 1991), such mutants are believed to be genuine 'hotspots' for mutation. However it is now becoming clear that the high number of repeat observations at some CG doublets (*e.g.* 30 examples at 31,008) are caused, at least in part, by a founder effect. A founder effect

is now well established (Thompson *et al*, 1990; Bottema *et al*, 1990b) for a mutation at a residue other than CG's (*e.g.* 27 examples at nucleotide 31,311) and there are many examples in the database of mutations repeating 2, 3 or occasionally 4 times. Most, but probably not all, of these will have a common origin.

A new feature this year is the inclusion of information in the comments section, on new (or de novo) mutations by the UK and German coordinators. In addition, we note whether the mutation occurred in the mother or in the maternal grandfather or grandmother. The German coordinator has also included the age of the parent of the child carrying, or affected by, the mutant gene. Because only the UK and German coordinators list new mutations, and even their data is incomplete, this database cannot yet be usefully analysed for their frequency.

The distribution of mutants according to protein domains and control regions within the gene (see Summary Table) shows that mutations have been detected in all categories listed except the poly(A) site. Remarkably, there are now 11 molecularly unique mutants occurring within a short region of the promoter, and these are invaluable in studying gene regulation (Crossley & Brownlee, 1990). Missense mutations within exons give valuable information as to the importance of particular amino acid residues. For example, it is reassuring to note that 3 different mutations have been discovered at the active site serine (amino acid 365) as well as 1 at the proposed active site aspartate (amino acid 269), although none is yet known for the proposed active site histidine at amino acid 221. Mutations at 6 of the 12 γ carboxyglutamyl residues have now been detected, confirming their critical role for the function of factor IX.

The second (1991) edition of this data base (Giannelli *et al* 1991) is now available from the EMBL file server. It can be obtained by sending the command GET HAEMB.DAT to NETSERV@EMBL-HEIDELBERG.DE (Internet address). A documentation file introducing the format used, which differs from this present data base because of restrictions imposed in transmitting data by electronic mail, is obtained by sending GET HAEMB.DOC to the same address. The format used in the 1991 File Server version of the data base is somewhat related to that used for the familiar EMBL nucleic acid sequence data base and should allow easy computer searching for particular features. The new entries in this 1992 edition of the data base will be reformatted and transferred to the EMBL file server during 1992.

The data base was compiled from separate lists updating the previous year's list prepared by coordinators for the different countries as follows:- Giannelli and Green representing the UK, Sweden and Iceland; High and Sommer representing USA; Lillicrap representing Canada; Ludwig and Olek representing Germany; Reitsma representing The Netherlands; Goossens representing France; Yoshioka representing Japan; and Brownlee, the rest of the world and central coordinator. New data or notification of errors or omissions should be sent to the individual country coordinators. This database is available from individual country coordinators on floppy discs written in Wordperfect 5.1 on an IBM PS2 computer.

ACKNOWLEDGEMENT

We thank Suzanne Motyka for acting as coordinating secretary.

REFERENCES

Alkan M., Rodriguez Ponte M., Malik N.J. et al (1991) Nucl.Acids Res. 19: 409

Anson D.S., Choo K.H., Rees D.J.G. et al (1984) EMBO J. 3 1054–1064

Attree O., Vidaud D., Vidaud M., Amselem S., Lavergne J.M. and Goossens M. (1989) Genomics 4: 266–272

Bajaj S.P., Spitzer S.G., Welsh W.J., Warn–Cramer B.J., Kasper C.K., and Birktoft J.J. (1990) J.Biol.Chem. 265: 2956–61.

Bentley A.K., Rees D.J.G., Rizza C. and Brownlee G.G. (1986) Cell 45: 343–348

Bertina R.M., Van der Linden I.K., Mannucci P.M. Reinalda-Poot H.H., Cupers R., Poort S.R. and Reitsma P.H. (1990) J.Biol.Chem. 265: 10876–10883

Bottema C.D.K., Koeberl D.D. and Sommer S.S. (1989a) Lancet (ii), 526–529

Bottema C.D.K., Ketterling R.P., Cho H.I. and Sommer S.S. (1989b) Nucl. Acids Res. 17: 10139.

Bottema C.D.K., Ketterling R.P., Koeberl D.D. et al (1990a) Nucl. Acids Res. 18: 1924

Bottema C.D.K., Koeberl D.D., Ketterling R.P., Bowie E.J.W., Taylor S.A.M., Lillicrap D., Shapiro A., Gilchrist G., and Sommer S.S. (1990b) Br.J.Haematol. 75: 212–216.

Bottema C.D.K., Ketterling R.P., Yoon H.S., and Sommer S.S. (1990c) Am.J.Hum.Genet. 47: 835–841.

Bottema C.D.K., Ketterling R.P., Ii S., Yoon, H.S., Phillips J.A. and Sommer S.S. (1991a) Am. J. Hum. Genet. 49: 820–838.

Bottema, C.D.K., Bottema M.J., Ketterling R.P., Yoon H.S., Janco R.L., Phillips J.A. and Sommer S.S. (1991b) Am. J. Hum. Genet. 49: 839–850.

Brownlee G.G. (1988) in: Recent Advances in Haematology, ed. Hoffbrand A.V. (Churchill Livingstone) vol.5, pp. 251–264

Chan V. et al, Dept of Medicine, Univ. of Hong Kong, Queen Mary Hospital, Hong Kong, unpublished

Chan V., Yip B., Tong T.M.F. et al (1991) Br. J. Haematol. 79: 63–69.

Chen S.H., Scott C.R., Lovrien E.W. et al (1989a) Am. J. Hum. Genetics 44: 567–569

Chen S.H., Thompson A.R. Zhang M., and Scott C.R. (1989b) J. Clin. Invest. 84: 113–118

Chen S.H., Zhang M., Lovrien E.W., Scott C.R. and Thompson A.R. (1991) Hum. Genet. 87: 177–182.

Chen S.H. and Thompson A.R. Unpublished observation, Dept. of Pediatrics, Univ. of Washington, and Puget Sound Blood Center, Seattle, WA.

Crossley P.M., Winship P.R., Black A., Rizza C. and Brownlee G.G. (1989) Lancet (i): 60

Crossley P.M. and Brownlee G.G. (1990) Nature 345: 444–446

Crossley M., Winship P.R., Austen D.E.G., Rizza C.R. and Brownlee G.G. (1990) Nucl.Acids Res. 18: 4633

Davis L.M, McGraw R.A., Ware J.L. et al (1987) Blood 69: 140–143

De la Salle C., INSERM U311, CRTS, 10 Rue Spielman, 67085 Strasbourg, France, unpublished

Demers D.B., Odelberg S.J. and Fisher L. (1990) Nucl. Acids Res. 18: 5575.

Denton P.H., Fowlkes D.M., Lord S.T., and Reisner H.M. (1988) 72: 1407–1411

Diuguid D.L., Rabiet M.J., Furie B.C. et al (1986) Proc. Natl. Acad. Sci. (USA) 83: 5803–5807

Diuguid D.L., Rabiet M.J., Furie B.C., and Furie B. (1989) Blood 74: 193–200

Driscoll M.C., Bouhassira E. and Aledort L.M. (1989a) Blood 74: 737–742

Driscoll M.C., Aledort L.M., Hilgartner M.W. (1989b) Blood 74: 254a (abstract) and unpublished observation.

Fraser B.M., Poon M-C. and Hoar D.I. (1992) Hum. Genet. in press

Freedenberg D.L. and Black B. (1991) Thromb. Haemost. 65: 964 (abstract)

Freedenberg D.L., Chen S.H., and Scott R. (1989) Am J. Hum. Genetics 45 (suppl): A186 (abstract)

Geddes V.A., LeBonniec B.F., Louie G.V., Brayer G.D., Thompson A.R., and MacGillivray R.T.A. (1989) J.Biol.Chem. 264: 4689–4697

Ghanem N., et al INSERM U91, C.H.U. Henri Mondor, Creteil, France, unpublished.

Giannelli F., (1989) in: Baillière's Clinical Haematology, ed. Tuddenham E.G.D. (Baillière & Tindall, London) vol.2, pp.821–848

Giannelli F., Green P.M., High K.A. et al (1991) Nucl.Acids Res. 19(suppl): 2193–2219

Gispert S., Vidaud M., Vidaud D., Gazengel C., Boneu B. and Goossens M. (1989) Am.J.Hum.Genet. 45 Abstract 739

Green P.M., Bentley D.R., Mibashan R.S., Nilsson I.M. and Giannelli F. (1989) EMBO J. 8: 1067–1072

Green P.M., Montandon A.J., Bentley D.R., Ljung R., Nilsson I.M. and Giannelli F. (1990) Nucl. Acids Res. 18: 3227–3231

Green P.M., Montandon A.J., Bentley D.R. & Giannelli F. (1991a) Blood Coagulation & Fibrinolysis 2: 539–565

Green P.M., Montandon A.J., Ljung R., Bentley D.R., Kling S.R., Nilsson I.M. and Giannelli F. (1991b) Br. J. Haematol. 78: 390–397

Green P.M., Montandon A.J., Ljung R., Nilsson I.M. and Giannelli F. (1992) Thromb. Haemost. in press.

Green P.M., Montandon A.J. et al, Paediatric Research Unit, Guy's Hospital Tower, London, UK, unpublished

Green P.M., Saad et al, Paedeatric Research Unit, Guy's Hospital Tower, London, UK, unpublished

Hamaguchi M., Matsushita T., Tanimoto M. et al (1991) Thromb. Haemost. 65: 514–520

Handford P.A. et al (1991) Nature 351: 164–167

Hirosawa S., Fahner J.B., Salier J.P., Wu C.T., Lovrien E.W. and Kurachi K. (1990) Proc. Natl.Acad.Sci.USA 87: 4421–4425.

Hougie C. and Twomey J.J. (1967) Lancet i: 699

Huang M.N., Kasper C.K., Roberts H.R, et al (1989) Blood 73: 718–721

Jagadeeswaran P., Dept of Cellular & Structural Biology, Univ. of Texas, San Antonio, Texas, USA, unpublished

Ketterling R.P., Bottema C.D.K., Koeberl D.D., Li, S. and Sommer, S.S. (1991a) Hum.Genet. 87: 333–337

Ketterling R.P., Bottema C.D.K., Phillips J.A. and Sommer S.S. (1991b) Genomics 10: 1093–1096

Koeberl D.D., Bottema C.D.K., Buerstedde J.M., and Sommer S.S. (1989) Am. J. Hum. Genetics 45: 448–457

Koeberl D.D., Bottema C.D.K., Ketterling R.P., Bridge P.J., Lillicrap D.P., and Sommer S.S. (1990a) Am.J.Hum.Genet. 47: 202–217.

Koeberl D.D., Bottema C.D.K., Sarkar G., Ketterling R.P., Chen S.H. and Sommer S.S. (1990b) Hum.Genet. 84: 387–390.

Liddell M.B., Lillicrap D.P., Peake I.R. and Bloom A.L. (1989a) Br.J.Haematol. 72: 208–215

Liddell M.B., Peake I.R., Taylor S.A.M. et al (1989b) Br.J.Haematol 72. 556–560

Lillicrap D., Windsor S. et al, Dept. Pathology, Queen's University, Kingston, Ont. Canada, unpublished

Lin S.-W. and Shen M.C. (1991) Thromb. Haemost. 66: 459–463

Ljung R. et al, Dept. Coagulation Disorders, Univ. Lund, Malmö General Hospital, Malmö, Sweden, unpublished

Lozier J.N., Monroe D.M., Stanfield-Oakley S.A. et al (1990) Blood 75: 1097–1104

Ludwig M., Schwaab R., Olek K., Brackmann H.H., Egli H. (1988) Thromb. Haemost. 59: 340.

Ludwig M., Schwaab R., Eigel A., Horst J., Egli H., Brackmann H-H and Olek K. (1989) Am.J.Hum.Genet. 45: 115–122

Ludwig M., Brackmann H.H. and Olek K. (1991) Klin.Wochenschr. 69: 196–200

Ludwig M., Grimm T., Brackmann H.H., Olek K. (1992a) Am. J. Hum. Genet. (in press)

Ludwig M., Sabharwaj A.K., Brackmann H.H., Olek K., Smith K.J., Birktoft J.J., Bajaj S.P. (1992b) Blood (in press)

Ludwig M. et al, Inst. Exp. Hämatologie, Univ. of Bonn, 5300 Bonn 1, W.Germany, unpublished

Maekawa H., Sugo T., Muramatsu S. et al (1991) Jpn. J. Thromb. Haemost. 2: 400 (abstract)

Matsushita T., Tanimoto M., Yamamoto K., Sugiura I., Hamaguchi M., Takamatsu J., Kamiya T. and Saito H. (1990) J.Lab.Clin.Med. 116: 492–497

Miyata T., Sakai T., Sugimoto M., Naka H., Yamamoto K., Yoshioka A., Fukui H., Mitsui K., Kamiya K., Umeyama H. and Iwanaga S. (1991) Biochemistry 30: 11286–11291

Monroe D.M., McCord D.M., Huang M.N. et al (1989) Blood 73: 1540–1544

Montandon, A.J., Green, P.M., Giannelli, F. and Bentley, D.R. (1989) Nucl. Acids Res. 17: 3347–3358

Montandon, A.J., Green, P.M., Bentley, D.R., Ljung, R., Nilsson, I.M. and Giannelli, F. (1990a) Hum. Gen. 85: 200–204

Montandon A.J., Makris M., Green P.M., Coffey A.J., Preston F.E. and Giannelli F. (1990b) Br.J.Haematol. 76: suppl.1 p.8

Nishimura T., Naka H., Kuze K. et al (1990) Acta Haematol. Jpn. 53: 1030–1035

Nishimura H., Takeya H., Suehiro K., Okamura T., Murakawa M., Niho Y., Miyata T. and Iwanaga S. (1991) Thromb. Haemost. 65: 712 (abstract)

Noyes C.M., Griffith M.J., Roberts H.R., and Lundblad R.L. (1983) Proc. Natl. Acad. Sci. (USA) 80: 4200–4202

Pang C.P., Crossley M., Kent G. and Brownlee G.G. (1990) Nucl.Acids Res. 18: 6731–6732

Picketts et al, Department of Pathology, Queen's University, Kingston, Ontario, Canada, unpublished

Picketts D.J., Bridge P.J. and Lillicrap D. (1990) Blood 76 (Suppl.1): 1721a (abstract)

Poon M.-C., Chui D.H.K., Patterson M., Starozik D.M., Dimnik L.S. and Hoar D.I. (1987) J.Clin.Invest. 79: 1204–1209

Poon M.-C. and Sinclair G.D. Dept. Haematology, Univ. Calgary, Alberta, Canada, unpublished

Poort S.R., Briët E., Bertina R.M. and Reitsma P.H. (1989a) Nucl. Acids Res. 17: 3614

Poort S.R., Briët E., Bertina R.M. and Reitsma P.H. (1989b) Nucl. Acids. Res. 17: 5869

Poort S.R., Briët E., Bertina R.M. and Reitsma P.H. (1990) Thromb.Haemost. 64: 379–384

Poort S.R. et al, Hemostasis & Thrombosis Research Unit, 2300 RC Leiden, The Netherlands, unpublished.

Rao K.J., Lyman G., Hamsabhushami K. et al (1990) Molecular & Cellular Probes, in press

Rees D.J.G., Rizza C.R. and Brownlee G.G. (1985) Nature 316: 643–645

Reitsma P.H., Bertina R.M., Ploos van Amstel J.K., Riemens A. and Briët E. (1988) Blood 72: 1074–1076

Reitsma P.H., Mandalaki T, Kasper C.K. and Briët E. (1989) Blood 73: 743–746

Reitsma P.H. et al, Hemostasis & Thrombosis Research Unit, 2300 RC Leiden, The Netherlands, unpublished

Ritchie D.B.C., Tam B.M. and MacGillivray R.T.A. (1989) Alberta Heritage Foundation for Medical Research Heritage Days. Abstract

Ritchie D.B.C., Tam B.M. and MacGillivray R.T.A. Dept. Biochem. Univ. Alberta, Edmonton, Alberta, Canada, unpublished

Rose V. and High K.A. Unpublished observation, Depts. of Medicine and Pathology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC.

Royle G., Van de Water N.S., Berry E., Ockelford P.A. and Browett P.J. (1990) Brit.J.Haematol. 77: 191–194

Sakai T., Yoshioka A., Yamamoto K., Niinomi K., Fujimura Y., Fukui H., Miyata T. and Iwanaga S. (1989) J.Biochem. 105: 756–759

Schach B.G., Yoshitake S., Davie E.W. (1987) J. Clin. Invest. 80: 1023–1028

Siguret V., Amselem S., Vidaud M. et al (1988) Br.J.Haematol. 70: 411–416

Solera J., Magallón M., Martín-Villar J. and Coloma A. (1991) Br. J. Haematol. 78: 385–389

Solera J. et al Dept. Biochem. Univ. Autonoma, Madrid, 28029, Spain, unpublished

Spitzer S.G., Pendurthi U.R., Kasper C.K., and Bajaj S.P. (1988) J. Biol. Chem. 263: 10,545–10,548

Spitzer S.G., Warn-Cramer B.J., Kasper C.K., and Bajaj S.P. (1990a) Biochem. J. 265: 219–225

Spitzer S.G., Kuppuswamy, M.N., Saini R., Kasper C.K., Birktoft J.J. and Bajaj S.P. (1990b) Blood 76: 1530–1537.

Stanfield-Oakley S.A., Kaspar C. and High K.A., Univ. N. Carolina, Chapel Hill, N.C., USA, unpublished

Suehiro K., Kawabata S., Miyata T., Takeya H., Takamatsu J., Ogata K., Kamiya T., Saito H., Niho Y. and Iwanaga S. (1989) J.Biol.Chem. 264: 21257–21265

Suehiro K., Miyata T., Takeya H., Takamatsu J., Saito H., Murakawa M., Okamura T., Niho Y. and Iwanaga S. (1990) Thromb. Res. 60: 311–320

Sugimoto M., Miyata T., Kawabata S., Yoshioka A., Fukui H., Takahashi H. and Iwanaga S. (1988) J. Biochem. 104: 878–880

Sugimoto M., Miyata T., Kawabata S., Yoshioka A., Fukui H. and Iwanaga S. (1989) Br.J.Haematol. 72: 216–221

Tam B.M., MacGillivray R.T.A. and Ritchie D.B.C. (1991) Thromb. Haemost. 65: 968 (abstract)

Tarnower A. and Smith K., (1991) Thromb. Haemost. 65: 967 (abstract)

Taylor S.A.M. (1990). Ph.D. Thesis, Queen's University, Kingston, Ontario.

Taylor S.A.M., Liddell M.B., Peake I.R., Bloom A.L. and Lillicrap D.P. (1990) Br. J. Haematol. 75: 217–221.

Taylor S.A.M., Deugau K.V. and Lillicrap D.P. (1991) Proc.Natl.Acad.Sci.USA 88: 39–42

Thompson A.R., Chen S.H., and Brayer G.D. (1989) Blood 74 (suppl): 134a (abstract)

Thompson A.R. (1990) Prog.Hemost.Thromb. 10: 175–214

Thompson A.R., Bajaj S.P., Chen S.H. and MacGillivray R.T.A. (1990) Lancet i:418

Thompson A.R., Schoof J.M., Weinmann A.F. and Chen S.H. (1992) Thromb. Res. in press

Thompson A.R. and Chen S.H. (1992) Methods in Enzymology (Lorand & Mann, guest editors), in press

Toomey J.R., Stafford D., and Smith K. (1988) Blood 72 (suppl): 312a (abstract)

Tsang T.C., Bentley D.R., Mibashan R.S. and Giannelli F. (1988) EMBO J. 7: 3009–3015

Vidaud M., Attree O., Schaad O., Vidaud D., Edelstein S. and Goossens M. (1988) Blood 72 Suppl.1, Abstract 1161
Vidaud M. (1990) Thesis, Paris VII University
Vidaud M., National Institute of Health and Medical Research, 51 Av. du Maréchal de Lattre de Tassigny, 94010 Creteil, France, unpublished.
Wang N.S., Chen S.H. and Thompson A.R. (1990a) Thromb. Haemost. 64: 302–306
Wang N.S., Zhang N.S., Thompson A.R., and Chen S.H. (1990b) Thromb. Haemost. 63: 24–26
Ware J., Davis L., Frazier D. et al (1988) Blood 72: 820–822
Ware J., Diuguid D.L., Liebman H.A. et al (1989) J. Biol. Chem. 264: 11,401–11,406
Winship P.R. (1986) D.Phil thesis, Oxford University
Winship P.R. (1990) Nucl. Acids Res. 18: 1310
Winship P.R. and Dragon A.C. (1991) Br.J.Haematol. 77: 102–109
Yoshitake S., Schach B.G., Foster D.C., Davie E.W. and Kurachi K. (1985) Biochemistry 24: 3736–3750

Summary Table derived from main database

Location[1]	Exon	Nucleotide number[3]	Number of mutants[2]	Unique molecular events[2]
Signal peptide (−46 to −18)	a	30–116	4	4
Propeptide (−17 to −1)	b	6,326–6,375	41	5
Gla	b	6,376–6,489	50	32
(1 to 46)	c	6,678–6,701		
EGF (1st)	d	10,392–10,505	50	22
(47 to 84)				
EGF (2nd)	e	17,669–17,797	24	19
(85 to 127)				
activation	f	20,363–20,565	66	29
(128 to 195)				
catalytic	g	30,039–30,153	29	15
(196 to 415)	h	30,822–31,372	255	111
Subtotal			519	237
Promoter			16	11
Donor splice sites			20	15
Acceptor splice sites			13	11
Cryptic splice[4]			6	4
Poly(A) site			0	0
Totals			**574**	**278**

[1] Amino acid numbers used (Anson et al, 1984)
[2] Excluding normal variants within double mutants
[3] For numbering, see Yoshitake et al (1985)
[4] These are possible new splice sites within exons

Characterized Point Mutations and Short Deletions/Additions in patients with Haemophilia B (Christmas disease)

Patient[7]	Clotting (normal =100%)	Antigen (normal =100%)	Nucleotide[1,3,8] position & mutation	Amino acid[1] change	Comments[3]	Reference	Patient Identity Number[10]
HB5, Japan	<1	<1	-793, G→A	None	Double (see 20,551), N?	Matsushita et al (1990)	1
Brandenburg	<1	<1	-26, G→C	None	Promoter	Ludwig et al	395
Leyden 1	<1→60[5]	<1→60	-20, T→A	None	Promoter	Reitsma et al (1988)	2
Datteln	<1→36[5]	<1→36[5]	-20, T→A	None	Promoter	Ludwig et al	3
Marseille	9[5]		-20, T→C	None	Promoter, de novo in MGF	Ghanem et al	396
High Wycombe	13→70[5]		-6, G→A[4]	None	Promoter	Crossley et al (1990)	4
Leyden, USA1	10	13	-6, G→A[4]	None	Promoter	Hirosawa et al (1990)	5
Toulouse	1→30[5]	1→30[5]	-6, G→C	None	Promoter	Gispert et al (1989)	6
Toronto 20	3		-5, A→T	None	Promoter	Picketts et al	7
Leyden, USA2	<2→37[5]		6, T→A	None	Promoter	Freedenberg, Black (1991)	397
Leyden, NZ	1→32[5]		8, T→C	None	Promoter, C/EBP binding site	Royle et al (1990)	8
Leyden 2	<1→60[5]	<1→60[5]	13, A→G	None	Promoter	Reitsma et al (1989)	9
HB13	32		13, A→G	None	Promoter	Koeberl et al (1989)	10
Norwich	3→35[5]		13, A→G	None	Promoter	Crossley et al (1989)	11

Poitiers	12[5]		13, A→G	None	Promoter	Ghanem et al	398
Leyden 3	<1→60[5]	<1→60[5]	13, Δ1	None	Promoter	Reitsma et al (1989)	12
Aachen	<1		13, A→C	None	Promoter, C/EBP-binding site	Ludwig et al	13
Riegelsberg	<1	<1	37, G→A[4]	-44, R→H	Double (see 31,084), N	Ludwig et al (1992a)	14
Calgary 4	<1	<1	48, A→T	-40, I→F	Double (see 31,133), N	Poon & Sinclair	399
UK 22	2	<1	79, T→A	-30, I→N	Signal peptide	Green, Montandon et al	15
HB130	20		111, T→C	-19, C→R	Signal peptide	Bottema et al (1991b)	400
UK 36		<1	111-120, Δ10	-19	Frameshift, Donor splice (a)	Green, Montandon et al	16
Meaux	<1		112, Δ1	-19	Frameshift	Ghanem et al	401
Recklinghausen	<1	<1	114, Ins AT	-18	Frameshift	Ludwig et al	17
HB64	<1	2	117, G→A	-17, V→I	Donor splice (a)?	Bottema et al (1991a)	402
Malmö 33	3	2	122, G→A		Donor splice (a), de novo in MGF	Green et al (1991b)	18
HB135			6,320, T→G		Acceptor splice (b)	Bottema et al (1991b)	403
UK 70	<1		6,325, G→A		Acceptor splice (b)	Green, Saad et al	404
Spain	<1		6,325, G→T		Acceptor splice (b)	Ghanem et al	405
Boxtel	4	36	6,364, C→T[4]	-4, R→W		Reitsma et al	19
Heiden	<1		6,364, C→T[4]	-4, R→W		Ludwig et al	20
Lienen	<1	40	6,364, C→T[4]	-4, R→W		Ludwig et al	21

Malmö 6	<1	26-34	6,364, C→T[4]	-4, R→W		Green et al (1989)	22
Malmö 19	2-3	36	6,364, C→T[4]	-4, R→W		Green et al (1990)	23
Malmö 20	2	27	6,364, C→T[4]	-4, R→W	de novo in MGF or MGM	Green et al (1990)	24
Malmö 40	1	30	6,364, C→T[4]	-4, R→W		Green et al (1990)	25
Malmö 48	2		6,364, C→T[4]	-4, R→W		Ljung et al	406
UK 130	2		6,364, C→T[4]	-4, R→W		Green, Montandon et al	26
UK 134	<1		6,364, C→T[4]	-4, R→W		Green, Montandon et al	27
UK 129	4		6,364, C→T[4]	-4, R→W		Green, Montandon et al	28
Paris 2	<1	28	6,364, C→T[4]	-4, R→W		Ghanem et al	407
HB90	9	44	6,364, C→T[4]	-4, R→W		Bottema et al (1991a)	408
Bendorf	<1		6,365, G→T	-4, R→L		Ludwig et al	29
Beuten	<1	45	6,365, G→T	-4, R→L		Ludwig et al	30
Gleiwitz	<1	49	6,365, G→T	-4, R→L		Ludwig et al	31
Kingston 1 (HB55)	<1	27	6,365, G→T	-4, R→L		Koeberl et al (1990a)	32
Caen	2	52	6,365, G→T	-4, R→L		Ghanem et al	409
Besancon	3		6,365, G→T	-4, R→L		Ghanem et al	410
Oxford 3	<1	89	6,365, G→A[4]	-4, R→Q		Bentley et al (1986)	33
San Dimas	<1	98	6,365, G→A[4]	-4, R→Q	Abnormal carboxylation, circulates with propeptide	Ware et al (1989)	34
Dortmund	<1		6,365, G→A[4]	-4, R→Q		Ludwig et al	36

Hong Kong 2	4	93	6,365, G→A[4]	-4, R→Q		Chan et al (1991)	37
Hong Kong 3	1	82	6,365, G→A[4]	-4, R→Q		Chan et al (1991)	38
Kawachinagano	<1	46	6,365 G→A[4]	-4, R→Q		Sugimoto et al (1989)	39
Kingston 2 (HB56)	<1		6,365, G→A[4]	-4, R→Q	Patients ID=35 & 40 are the same	Koeberl et al (1990a) Bottema et al (1990a)	40 35
Münster	<1		6,365, G→A[4]	-4, R→Q		Ludwig et al	41
Strasbourg II	<1	40	6,365, G→A[4]	-4, R→Q		de la Salle	42
Troed-y-Rhiw	4	80	6,365, G→A[4]	-4, R→Q		Liddell et al (1989a)	43
UK 3	1	48	6,365, G→A[4]	-4, R→Q		Green et al (1989)	44
UK 4	1	45	6,365, G→A[4]	-4, R→Q		Green et al (1989)	45
Unnamed			6,365, G→A[4]	-4, R→Q		Ghanem et al	411
Taipei 1	<1	38	6,365, G→A[4]	-4, R→Q		Lin & Shen (1991)	412
NZ 2	2		6,365, G→A[4]	-4, R→Q		Van de Water et al	413
Seattle A	2	40	6,365, G→A[4]	-4, R→Q		Thompson et al (1992)	414
Seattle B			6,365, G→A[4]	-4, R→Q		Thompson et al (1992)	415
Seattle C	<1	65	6,365, G→A[4]	-4, R→Q		Thompson et al (1992)	416
Seattle D	<1	41	6,365, G→A[4]	-4, R→Q		Thompson et al (1992)	417
HB159	<1	37	6,365, G→A[4]	-4, R→Q	Haplotype differs from 35	Bottema et al (1991b)	418
Seattle E	<1	37	6,372, G→T	-2, K→N		Thompson et al (1992)	419

Cambridge	<1	80	6,375, G→C/T[2]	-1, R→S	Abnormal carboxylation, circulates with propeptide	Diuguid et al (1986)	46
London, Ont 1 (HB54)	6		6,379, A→G	2, N→D		Koeberl et al (1990a)	47
Autun	<1		6,379, Δ1	2	Frameshift	Ghanem et al	420
HB47	7		6,390, A→G		New donor splice in exon(b)	Koeberl et al (1990a)	389
UK 12	<1		6,392, Δ1	6	Frameshift, Inhibitor	Green et al (1989)	48
Oxford b2	5	5	6,395, A→C	7, E→A	Gla	Winship & Dragon (1991)	49
Malmö 8	<1	<0.1	6,398-9, Δ2	8	Frameshift, de novo in MGF or MGM	Green et al (1991b)	50
Tours	<1	<0.1	6,398-9, Δ2	8	Frameshift	Ghanem et al	421
HB151	2	45	6,398, A→G	8, E→G	Gla	Bottema et al (1991b)	422
UK 84	14	118	6,400, T→A	9, F→I		Green, Montandon et al	51
Madrid 3	<1	<1	6,401-10, Δ10	9	Frameshift, Inhibitor	Solera et al	423
Bonn 2	<1	<1	6,402-6, Δ5	9	Frameshift, Inhibitor de novo in mother(22)	Ludwig et al (1992a)	52
Oxford b3	<1	0.2	6,406, C→T	11, Q→Stop		Winship & Dragon (1991)	53
Hong Kong 1	3	97	6,410, G→C	12, G→A		Chan et al (1991)	54
Mühlheim/Ruhr	4	4	6,416-17, Δ2	14	Frameshift, somatic mosaic?	Ludwig et al	55
Unnamed	<1		6,424, G→A	17, E→K	Gla	Ghanem et al	424
Heessen	<1		6,427, T→C	18, C→R	de novo in mother(25)	Ludwig et al (1992a)	56

Zutphen	<1	100	6,427, T→C	18, C→R		Reitsma et al	57
HB116	<1		6,428, G→A	18, C→Y		Bottema et al (1991a)	425
Nagoya 4	<1	52	6,436, G→A	21, E→K	Gla	Hamaguchi et al (1991)	426
Jie Long	<1	35	6,442, T→C	23, C→R		Lin & Shen (1991)	427
UK 115	<1	19	6,443, G→A	23, C→Y		Green, Montandon, et al	58
Rheidt	2	32	6,449, T→C	25, F→S	de novo in MGF(27) or MGM(24)	Ludwig et al (1992a)	59
Seattle 3	<1	30	6,454, G→A	27, E→K	Gla	Chen et al (1989b)	60
Chongqing	<1	3	6,455, A→T	27, E→V	Gla	Wang et al (1990b)	61
HB28	<1	<1	6,460, C→T[4]	29, R→Stop		Koeberl et al (1990b)	62
HB61	<1	7	6,460, C→T[4]	29, R→Stop	Haplotype differs from 62	Koeberl et al (1990b)	63
Malmö 4	<1	<0.1	6,460, C→T[4]	29, R→Stop	Inhibitor, de novo in MGF or MGM	Green et al (1989)	64
UK 14	<1	2	6,460, C→T[4]	29, R→Stop		Montandon et al (1989)	65
UK 24	1		6,460, C→T[4]	29, R→Stop		Green et al (1990)	66
Unnamed	<1	<1	6,460, C→T[4]	29, R→Stop		Ludwig et al	67
UK 69	<1		6,460, C→T[4]	29, R→Stop		Green et al (1992)	68
UK 55	<1	<1	6,460, C→T[4]	29, R→Stop		Green et al (1992)	69
Taipei 2	<1	<6	6,460, C→T[4]	29, R→Stop		Lin & Shen (1991)	428
Taipei 3	<1	<6	6,460, C→T[4]	29, R→Stop		Lin & Shen (1991)	429

UK 195			6,460, C→T[4]	29, R→Stop		Green, Saad et al	430
Seattle F	<1	<1	6,460, C→T[4]	29, R→Stop	Inhibitor	Thompson et al (1992)	431
Pluvigner	19		6,461, G→A[4]	29, R→Q		Ghanem et al	432
NZ 3	18		6,461, G→A[4]	29, R→Q		Van de Water et al	433
Toronto 17 (HB58)	37		6,461, G→A[4]	29, R→Q	Double (see 30,134)	Koeberl et al (1990a)	70
HB2	30		6,461, G→A[4]	29, R→Q	Double (see 30,134)	Koeberl et al (1989)	71
Unnamed	20	70	6,461, G→A[4]	29, R→Q	Double (see 30,134)	Chen et al (1991)	72
Unnamed	10	77	6,461, G→A[4]	29, R→Q		Thompson & Chen (1992)	434
Malmö 9	<1	<0.1	6,466, Δ1	31	Frameshift	Green et al (1991b)	73
UK 180			6,472, Ins GG	33	Frameshift	Green, Saad et al	435
HB9	4		6,474, A→C	33, E→D	Gla	Koeberl et al (1989)	74
UK 10	<1	12	6,484-6, Δ3	37, ΔR	In frame	Green et al (1989)	75
Unnamed	<1		6,488, C→G	38, T→R		Ludwig et al	76
UK 89	7		6,488, C→T	38, T→I		Green, Saad et al	436
Ursem	1	<1	6,491-4 or 6,492-5, Δ4		Double (see 31,103), Donor splice (b) probably causes disease	Poort et al (1990)	77
Unnamed	<1	<1	6,491-4 or 6,492-5, Δ4		Donor splice (b)	Chen & Thompson	78
UK 83	<1	1	6,491-4 or 6,492-5, Δ4		Donor splice (b)	Green, Montandon et al	79

Paris 1			6,494, G→A		Donor splice (b)	Ghanem et al	437
HB74/77	1	1	6,495, T→C		Donor splice (b)	Bottema et al (1990c)	80
HB131	<1	<1	6,575, C→G		Triple (see 10,512 and 30,897), N	Bottema et al (1991b)	438
Malmö 10	<1	0.2	6,666-75, Δ10		Acceptor splice (c), de novo in MGF or MGM	Green et al (1991b)	81
HB7, Japan	<1	<1	6,680-1, Δ2	39	Frameshift, Inhibitor	Matsushita et al (1990)	82
Hoogeveen	14	96	6,690, A→G	43, K→E		Reitsma et al	83
Malmö 52	<1		6,693, C→T	44, Q→Stop		Ljung et al	439
UK 25	<1	<1	6,702, G→A	47, D→N	Donor splice (c)?	Green, Montandon et al	84
Oxford 2	0.5	0.4	6,704, T→G		Donor splice (c)	Winship (1986)	85
Pirmasens	<1	<1	6,704, T→C		Donor splice (c)	Ludwig et al	86
Nantes	3,13		6,704, T→C		Donor splice (c), female to female transmission	Ghanem et al	440
Toronto 16 (HB53)	3		10,391, G→A		Acceptor splice (d)	Koeberl et al (1990a)	87
Alabama	10	100	10,392, A→G	47, D→G		Davis et al (1987)	88
HB75	14	80	10,393, T→A	47, D→E		Bottema et al (1990c)	89
Oxford d3	1	90	10,393, T→A	47, D→E		Winship & Dragon (1991)	90
Tainan	28	101	10,394, G→A	48, G→R		Lin & Shen (1991)	441
Malmö 27	19	108	10,395, G→T	48, G→V	de novo in mother	Green et al (1991b)	91
UK 86			10,397-9, Δ3, Ins A	49	Frameshift	Green, Montandon et al	92

New London	<1	114	10,401, A→C	50, Q→P	Decreased XIa activation	Lozier et al (1990)	93
HB97	<1	20	10,406, G→T	52, E→Stop		Bottema et al (1991a)	442
Hollywood	11	58	10,415, C→G	55, P→A		Spitzer et al (1990b)	94
UK 7	10-12	49	10,415, C→G	55, P→A		Green et al (1989)	95
Unnamed	10	<30	10,415, C→G	55, P→A		Chen et al (1991)	96
Malmö 21	12	52	10,415, C→T	55, P→S		Green et al (1991b)	97
Malmö 22	26		10,416, C→T	55, P→L		Green et al (1991b)	98
Basel	<1		10,418, T→C	56, C→R		Alkan et al (1991)	99
Kleve	<1	<1	10,419, G→C	56, C→S		Ludwig et al (1991)	100
Toronto 2 (HB32)	1	2	10,419, G→A	56, C→Y		Koeberl et al (1990a)	101
UK 160	18		10,427, G→A	59, G→S		Green, Saad et al	443
HB141	2		10,428, G→T	59, G→V		Bottema et al (1991b)	444
Durham	14		10,430, G→A[4]	60, G→S		Denton et al (1988)	102
Kingston 3 (HB57)	10	11	10,430, G→A[4]	60, G→S	Same haplotype as 390	Koeberl et al (1990a)	103
Lelystad	13	34	10,430, G→A[4]	60, G→S		Poort et al	104
Oud en Nieuw Gastel	12	31	10,430, G→A[4]	60, G→S		Poort et al (1989b)	105
Purmerend	17	22	10,430, G→A[4]	60, G→S		Poort et al	106
UK 27	10		10,430, G→A[4]	60, G→S		Green et al (1990)	107
Unnamed	11	19	10,430, G→A[4]	60, G→S		Chen et al (1989b)	108

Unnamed	11	19	10,430, G→A[4]	60, G→S		Chen et al (1989b)	108
Unnamed	17	30	10,430, G→A[4]	60, G→S		Chen et al (1989b)	109
Unnamed	10	18	10,430, G→A[4]	60, G→S		Chen et al (1991)	110
Unnamed	15	54	10,430, G→A[4]	60, G→S		Chen & Thompson	111
Oxford d2	10	28	10,430, G→A[4]	60, G→S		Winship & Dragon (1991)	112
HB3/7	14	35	10,430, G→A[4]	60, G→S		Koeberl et al (1989)	390
HB4	16	41	10,430, G→A[4]	60, G→S		Koeberl et al (1989)	391
Unnamed	13	30	10,430, G→A[4]	60, G→S		Thompson & Chen (1992)	445
HB99	17		10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	446
HB103	12	32	10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	447
HB117	5		10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	448
HB133	10		10,430, G→A[4]	60, G→S	Haplotype differs from 390	Ketterling et al (1991b)	449
HB140	15		10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	450
HB146	22	23	10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	451
HB155	9	20	10,430, G→A[4]	60, G→S	Haplotype differs from 390	Ketterling et al (1991b)	452
HB160	15	20	10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	453
HB176	16		10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	454
HB184	9		10,430, G→A[4]	60, G→S	Same haplotype as 390	Ketterling et al (1991b)	455
Seattle G	11	31	10,430, G→A[4]	60, G→S		Thompson et al (1992)	456

Seattle H	15	30	10,430, G→A[4]	60, G→S		Thompson et al (1992)	457
HB154	<1	6	10,430, G→C	60, G→R		Bottema et al (1991b)	458
Toronto 6 (HB37)	1	2	10,431, G→A	60, G→D		Koeberl et al (1990a)	113
Oxford d1	3	117	10,442, G→A	64, D→N	ß Hydroxyaspartate	Winship & Dragon (1991)	114
UK 6	8	87	10,443, A→G	64, D→G	β Hydroxyaspartate	Green et al (1989)	115
Trier	<1		10,458, A→G	69, Y→C		Ludwig et al	116
UK 67	6	12	10,479, G→T	76, G→V	Female	Green, Montandon et al	117
UK 19			10,482, T→G	77, F→Y		Green, Montandon et al	118
UK 132			10,507-10, Δ4		Donor splice (d), normal carrier female	Green, Montandon et al	119
Toronto 8 (HB39)	2	3	10,512, A→G		Double (see 30,864),N?	Koeberl et al (1990a)	120
HB131	<1	<1	10,512, A→G		Triple (see 6575 & 30,897), N?	Bottema et al (1991b)	438
HB6	20		17,660-3, Δ 4		Acceptor splice (e)	Koeberl et al (1989)	121
Toronto 14 (HB48)	3	3	17,667, A→G		Acceptor splice (e)	Koeberl et al (1990a)	122
Malmö 11	<1	<1	17,668, G→C		Acceptor splice (e), de novo in MGF	Green et al (1991b)	123
Seattle 2	<1	<1	17,669, Δ1	85	Frameshift	Schach et al (1987)	124
Königswinter	<1		17,678, G→C	88, C→S		Ludwig et al	125
Fukuoka	2	66	17,689, A→C[2]	92, N→H		Nishimura et al (1991)	126
Chelles	<1		17,691, T→A	92, N→K		Ghanem et al	459

Unnamed	2		17,691, T→A	92, N→K		Ghanem et al	460
HB106	1	84	17,697, A→T	94, R→S		Bottema et al (1991b)	461
HB134	5		17,699, G→A	95, C→Y		Bottema et al (1991b)	462
HB157	<1	6	17,699, G→A	95, C→Y	Haplotype differs from 462	Bottema et al (1991b)	463
Edmonton 1	<1		17,700, C→A	95, C→Stop		Tam et al	127
Malmö 51	<1		17,700, C→G	95, C→W		Ljung et al	464
Hamilton 1 (HB45)	<1		17,710, T→C	99, C→R		Koeberl et al (1990a)	128
NZ 4	<1		17,718, Ins A	101	Frameshift	Van de Water et al	465
UK 50	<1	<1	17,727, Ins TT	105	Frameshift	Green, Montandon et al	129
Malmö 35	21	14	17,736, G→A	107, None	cryptic splice?	Green et al (1991b)	130
Malmö 42	20	24	17,736, G→A	107, None	cryptic splice?	Green et al (1991b)	131
Malmö 37	15	24	17,736, G→A	107, None	cryptic splice?	Green et al (1991b)	132
Unnamed	20	120	17,738, T→C	107, V→A		Chen et al (1991)	133
HB111	<1	<1	17,743, T→C	110, S→P		Bottema et al (1991a)	466
Leamington	13		17,756, G→C	114, G→A		Ritchie et al (1989)	134
Oxford e1	5	4	17,756, G→C	114, G→A		Winship & Dragon (1991)	135
Nastetten	<1		17,759, A→G	115, Y→C		Ludwig et al	136
UK 123	2	1	17,759, A→G	115, Y→C		Green, Saad et al	467
Malmö 7	<1	<0.1	17,761, C→T[4]	116, R→Stop	Double (see 30,890), de novo in MGF	Montandon et al (1990a)	137

HB149	<1	3	17,761, C→T[4]	116, R→Stop		Bottema et al (1991b)	468
UK 28	5	5	17,761, C→A	116, None	cryptic splice?	Green, Montandon et al	138
Würzburg	<1		17,764, Ins C	117	Frameshift, de novo in MGF(27) or MGM(21)	Ludwig et al (1992a)	139
UK 9	<1	0.4	17,773, A→T	120, N→Y		Green et al (1989)	140
HB88	2		17,786, G→A	124, C→Y		Bottema et al (1991a)	469
HB68	1		17,797, G→A	128, V→M	Donor splice (e)	Bottema et al (1991a)	470
Nörtingen	<1		17,798, G→T		Donor splice (e), de novo in MGM(27)	Ludwig et al (1992a)	141
Toronto 13 (HB44)	10		17,810, A→G		Donor splice (e), same haplotype as 143	Koeberl et al (1990a)	142
Toronto 15 (HB52)	10		17,810, A→G		Donor splice (e), same haplotype as 142	Koeberl et al (1990a)	143
UK 63			17,810, A→G		Donor splice (e)	Green, Montandon et al	144
Dakar	<1	<1	20,374, T→C	132, C→R		Vidaud (1990)	145
Malmö 12	<1	<0.1	20,375, G→T	132, C→F		Green et al (1991b)	146
HB115	<1		20,375, G→A	132, C→Y		Bottema et al (1991a)	471
Malmö 44			20,398, Δ1	140	Frameshift, de novo in MGF or MGM	Green et al (1991b)	147
Albuquerque	1	30	20,413, C→T[4]	145, R→C	Decreased XIa activation	Toomey et al (1988)	148
Cardiff 1	<1	66	20,413, C→T[4]	145, R→C		Liddell et al (1989b)	149
UK 21	1		20,413, C→T[4]	145, R→C		Green et al (1990)	150

UK 23	2	43	20,413, C→T[4]	145, R→C		Green et al (1990)	151
Oxford f1	3	54	20,413, C→T[4]	145, R→C		Winship & Dragon (1991)	152
UK 60	<1	13	20,413, C→T[4]	145, R→C		Green, Saad et al	472
UK 96	2		20,413, C→T[4]	145, R→C		Green, Saad et al	473
UK 125	<1		20,413, C→T[4]	145, R→C		Green, Saad et al	474
UK 194			20,413, C→T[4]	145, R→C		Green, Saad et al	475
Unnamed	2	40	20,413, C→T[4]	145, R→C		Thompson & Chen (1992)	476
Chapel Hill	8	100	20,414, G→A[4]	145, R→H	Decreased XIa activation	Noyes et al (1983)	153
Chicago 2	7	160	20,414, G→A[4]	145, R→H	Decreased XIa activation	Diuguid et al (1989)	154
HB25	4		20,414, G→A[4]	145, R→H		Koeberl et al (1989)	155
Malmö 17	4-11	91	20,414, G→A[4]	145, R→H		Green et al (1991b)	156
Malmö 23	7	148	20,414, G→A[4]	145, R→H		Green et al (1991b)	157
Malmö 32	5-8	115	20,414, G→A[4]	145, R→H		Green et al (1991b)	158
Malmö 36	7	110	20,414, G→A[4]	145, R→H		Green et al (1991b)	159
Malmö 38			20,414, G→A[4]	145, R→H		Green et al (1991)	160
Nagoya 3	4	100	20,414, G→A[4]	145, R→H		Suehiro et al (1990) Hamaguchi et al (1991)	161
HB76	6	100	20,414, G→A[4]	145, R→H	Haplotype differs from 155	Bottema et al (1990c)	162
NZ 5	3		20,414, G→A[4]	145, R→H		Van de Water et al	477
HB120	11		20,414, G→A[4]	145, R→H	Same haplotype as 155	Bottema et al (1991a)	478

HB156	4		20,414, G→A[4]	145, R→H	Same haplotype as 155	Bottema et al (1991b)	479
Toronto 21	7	60	20,414, G→T	145, R→L	Also 9,786, A→C in intron	Picketts et al (1990)	480
HB23	<1		20,466-78, Δ13	162	Frameshift	Koeberl et al (1989)	163
HB17	<1		20,497, C→T	173, Q→Stop		Koeberl et al (1989)	164
HB78	2	<1	20,501, Δ1	174	Frameshift	Bottema et al (1990c)	165
Malmö 13	<1	0.5	20,510, Δ1	177	Frameshift	Green et al (1991b)	166
Brest	<1	85	20,512, T→C	178, F→L	Double (see 20,518), N?	Vidaud	167
BM Nagoya 1	<1	100	20,518, C→T[4]	180, R→W	B_m	Suehiro et al (1989)	168
BM Nagoya 2	<1	120	20,518, C→T[4]	180, R→W	B_m	Hamaguchi et al (1991)	481
Deventer	<1	130	20,518, C→T[4]	180, R→W	B_m	Bertina et al (1990)	169
Dernbach	<1	69	20,518, C→T[4]	180, R→W	de novo in MGF(43)	Ludwig et al(1992a)	170
Idaho	<1		20,518, C→T[4]	180, R→W		Demers et al (1990)	171
New York 2	<1	74	20,518, C→T[4]	180, R→W	B_m	Driscoll et al (1989b)	172
Dominican Rep.	<1	95	20,518, C→T[4]	180, R→W	B_m	Driscoll et al (1989b)	173
Unnamed	<1		20,518, C→T[4]	180, R→W		Ghanem et al	482
Unnamed	<1	50	20,518, C→T[4]	180, R→W		Thompson & Chen (1992)	483
Unnamed	<1	110	20,518, C→T[4]	180, R→W		Thompson & Chen (1992)	484
Brest	<1	85	20,518, C→G	180, R→G	Double (see 20,512)	Vidaud	167
Madrid	<1	90	20,518, C→G	180, R→G	B_m	Solera et al (1991)	174

Hilo	<1	120	20,519, G→A[4]	180, R→Q	B_m, decreased XIa activation	Huang et al (1989) Monroe et al (1989)	175
Hilo, Fr	<1	120	20,519, G→A[4]	180, R→Q	B_m^6	Vidaud (1990)	176
Novara	<1	112	20,519, G→A[4]	180, R→Q	B_m	Bertina et al (1990)	177
Rheine 2	<1		20,519, G→A[4]	180, R→Q		Ludwig et al	178
Altenhunden	<1		20,519, G→A[4]	180, R→Q		Ludwig et al	179
HB79	3	100	20,519, G→A[4]	180, R→Q		Bottema et al (1990c)	180
Seattle I	<1	100	20,519, G→A[4]	180, R→Q		Thompson et al (1992)	485
Creston 1	<2		20,519, G→C	180, R→P		Ritchie et al	486
Milano	<1	130	20,521, G→T	181, V→F	B_m	Bertina et al (1990)	181
Cardiff 2	15	132	20,524, G→C	182, V→L	B_m^6	Taylor et al (1990)	182
Kashihara	<1	120	20,524, G→T	182, V→F	B_m^6	Sakai et al (1989) Nishimura et al (1990)	183
Tokyo	23	100	20,525, T→C	182, V→A	B_m^6	Maekawa et al (1991)	487
UK 54	<1	97	20,527-9, Δ3	183, ΔG	In frame	Green, Montandon et al	184
Bottrop 1	<1	92	20,531-3, Δ3	184, ΔG	In frame	Ludwig et al	185
HB5 Japan	<1	<1	20,551, C→T	191, Q→Stop	Inhibitor, Double (see -793)	Matsushita et al (1990)	1
Unnamed	<1	<1	20,551, C→A	191, Q→K		Chen et al (1991)	186
Seattle J	3	2	20,553, A→G	191, None	Cryptic splice	Thompson et al (1992)	488
Seattle K	4	4	20,560, T→C	196, W→R		Thompson et al (1992)	489

Seattle L	<1	1	20,560, T→A	196, W→R		Thompson et al (1992)	490
Malmö 5	<1	<0.1	20,561, G→A	194, W→Stop	Inhibitor	Green et al (1989)	187
UK 114			20,561, G→A	194, W→Stop	Normal carrier female	Green, Montandon et al	188
UK 44	<1	2	20,561, G→A	194, W→Stop		Green, Saad et al	491
Unnamed	<1	<1	20,564, A→G	195, Q→R		Ludwig et al	189
Oxford 1	<0.5	0.3	20,566, G→T		Donor splice (f)	Rees et al (1985)	190
Rotenburg	<1	<1	20,566, G→A		Donor splice (f)	Ludwig et al	191
UK 171	<1		30,038, G→A		Acceptor splice (g)	Green, Saad et al	492
HB102	<1	<1	30,038, G→C		Acceptor splice (g)	Bottema et al (1991)	493
UK 192	2		30,046, T→G	198, L→W		Green, Saad et al	494
Toronto 19	<1	<1	30,070, G→C	206, C→S		Taylor et al (1990)	192
HB142	<1		30,070, G→A	206, C→Y		Bottema et al (1991)	495
UK 43	<1		30,072, G→A	207, G→R		Green, Montandon et al	193
UK 37		1	30,076, G→A	208, G→D		Green, Montandon et al	194
Wültschkau	7		30,084, G→T	211, V→F		Ludwig et al	195
Unnamed	4		30,084, G→T	211, V→F		Chen et al (1991)	196
Bottrop 2	<1	<1	30,090, G→T	213, E→Stop		Ludwig et al	197
HB72	4	<1	30,096, T→C	215, W→R		Bottema et al (1990c)	198
Newcastle 1	4		30,096, T→C	215, W→R		Lillicrap et al	496

Unnamed	<1	0.1	30,097, G→A	215, W→Stop		Chen et al (1991)	199
Unnamed	4		30,100, T→C	216, I→T		Chen et al (1991)	200
HB41	7		30,100, T→C	216, I→T		Koeberl et al (1990a)	201
Malmö 39	4	4	30,100, T→C	216, I→T		Green et al (1991b)	202
HB65	15	4	30,101, T→G	216, I→M		Bottema et al (1991a)	497
Toronto 5 (HB36)	<1	3	30,112, C→T	220, A→V		Koeberl et al (1990a)	203
Toronto 11 (HB49)	4		30,112, C→T	220, A→V		Koeberl et al (1990a)	204
UK 93	2		30,117, T→A	222, C→S		Green, Saad et al	498
HB24	1		30,119, T→G	222, C→W		Koeberl et al (1989)	205
Toronto 17 (HB58)	37		30,134, T→C	None	Double, (see 6,461), N	Koeberl et al (1990a)	70
HB2	30		30,134, T→C	None	Double (see 6,461), N	Koeberl et al (1989)	71
Unnamed	20	70	30,134, T→C	None	Double (see 6,461), N	Chen et al (1991)	72
HB1	12		30,150, G→A[4]	233, A→T		Koeberl et al (1989)	206
Malmö 28	22		30,150, G→A[4]	233, A→T		Green et al (1991b)	207
Malmö 29	5-22	12	30,150, G→A[4]	233, A→T		Green et al (1991b)	208
Malmö 30	8-15		30,150, G→A[4]	233, A→T		Green et al (1991b)	209
Malmö 31	11	15	30,150, G→A[4]	233, A→T		Green et al (1991b)	210
Malmö 46	16		30,150, G→A[4]	233, A→T		Ljung et al	499
Opladen	10	13	30,150, G→A[4]	233, A→T		Ludwig et al	211

Edmonton 3	15		30,150, G→A[4]	233, A→T		Tam et al (1991)	212
Unnamed	10	11	30,150, G→A[4]	233, A→T		Chen et al (1991)	213
Unnamed	15	15	30,150, G→A[4]	233, A→T		Chen et al (1991)	214
Iceland 1	3	119	30,800, Ins A	None	Double (see 31,119), N	Green et al (1992)	215
HB6, Japan	<1	<1	30,821, G→A		Inhibitor, Acceptor splice (h)	Matsushita et al (1990)	216
Unnamed	<1	<1	30,821, G→A		Acceptor splice (h)	Chen et al (1991)	217
HB114	<1		30,821, G→A		Acceptor splice (h)	Bottema et al (1991a)	500
Spijkenisse	2	74	30,854, G→A	245, E→K		Reitsma et al	218
Monschau	3	39	30,855, A→T	245, E→V		Ludwig et al (1992b)	219
UK 140	<1		30,857, Δ1	246	Frameshift	Green, Saad et al	501
HB60	<1	<1	30,863, C→T[4]	248, R→Stop		Koeberl et al (1990a)	220
Malmö 3	<1	<0.1	30,863, C→T[4]	248, R→Stop	Inhibitor	Green et al (1989)	221
Malmö 14	<1	<0.1	30,863, C→T[4]	248, R→Stop	de novo in MGF	Green et al (1991b)	222
Malmö 15	<1	<0.1	30,863, C→T[4]	248, R→Stop	de novo in MGF	Green et al (1991b)	223
UK 26	<1	4	30,863, C→T[4]	248, R→Stop		Green et al (1990)	224
UK 47	<1		30,863, C→T[4]	248, R→Stop		Green et al (1992)	225
Unnamed	<1	<1	30,863, C→T[4]	248, R→Stop		Wang et al (1990a)	226
Las Cruces	<1	<1	30,863, C→T[4]	248, R→Stop		Tarnower & Smith (1991)	227
Artesia	<1	<1	30,863, C→T[4]	248, R→Stop		Tarnower & Smith (1991)	228

Unnamed	<1	<1	30,863, C→T[4]	248, R→Stop		Ludwig et al	229
Calgary 7	<1	<1	30,863, C→T[4]	248, R→Stop		Poon & Sinclair	502
HB136			30,863, C→T[4]	248, R→Stop	Haplotype differs from 220	Bottema et al (1991b)	503
Albuquerque 4	<1	<1	30,863, C→T[4]	248, R→Stop		Tarnower & Smith (1991)	504
HB144	<1		30,863, C→T[4]	248, R→Stop	Same haplotype as 220	Bottema et al (1991b)	505
Seattle 4	3	4	30,864, G→A[4]	248, R→Q		Chen et al (1989b)	230
Toronto 8 (HB39)	2	3	30,864, G→A[4]	248, R→Q	Double (see 10,512)	Koeberl et al (1990a)	120
Unnamed	1	1	30,864, G→A[4]	248, R→Q		Chen et al (1991)	231
Unnamed	4	3	30,864, G→A[4]	248, R→Q		Chen et al (1989b)	232
Dreihacken	3	4	30,864, G→A[4]	248, R→Q		Ludwig et al (1992b)	233
HB59	3	2	30,864, G→A[4]	248, R→Q		Koeberl et al (1990)	392
UK 118	<1		30,864, G→A[4]	248, R→Q		Green, Saad et al	506
UK 148			30,864, G→A[4]	248, R→Q		Green, Saad et al	507
Calgary 5	3	<10	30,864, G→A[4]	248, R→Q		Poon & Sinclair	508
Calgary 6	6	9	30,864, G→A[4]	248, R→Q		Poon & Sinclair	509
HB98	6	3	30,864, G→A[4]	248, R→Q	Same haplotype as 120	Bottema et al (1991a)	510
HB104	15		30,864, G→A[4]	248, R→Q	Haplotype differs from 120	Bottema et al (1991a)	511
Unnamed	3	3	30,864, G→A[4]	248, R→Q		Thompson & Chen (1992)	512
Leiria	<1	<1	30,875, C→T[4]	252, R→Stop		Siguret et al (1988)	234

Malmö 41	<1	6	30,875, C→T[4]	252, R→Stop	de novo in mother	Green et al (1991b)	235
Portland	<1	<1	30,875, C→T[4]	252, R→Stop		Chen et al (1989a)	236
Toronto 18	3		30,875, C→T[4]	252, R→Stop		Taylor (1990)	237
Calgary 1	<1	<1	30,875, C→T[4]	252, R→Stop		Poon et al (1987)	238
Hong Kong 8	<1	<10	30,875, C→T[4]	252, R→Stop		Chan et al	239
UK 120	<1		30,875, C→T[4]	252, R→Stop		Green et al (1992)	240
UK 149			30,875, C→T[4]	252, R→Stop		Green, Saad et al	513
HB91	3	<1	30,875, C→T[4]	252, R→Stop		Bottema et al (1991a)	514
UK 16	13		30,876, G→T	252, R→L		Green, Montandon et al	241
Malmö 7	<1	<0.1	30,890, C→T	257, H→Y	Double (see 17,761), N	Montandon et al (1990a)	137
HB131	<1	<1	30,897, A→G	259, Y→C	Triple (see 6575 & 10,512)	Bottema et al (1991b)	438
HB8	24		30,900, A→G	260, N→S		Koeberl et al (1989)	242
UK 88	7		30,924, A→G	268, H→R		Green, Montandon et al	243
Malmö 50	<1		30,927, A→T	269, D→V	Proposed active site Asp	Ljung et al	515
UK 15	2	2	30,929, A→T	270, I→F		Green, Montandon et al	244
Beuren	<1	<1	30,930, T→C	270, I→T		Ludwig et al	245
HB96	<1	<1	30,930, T→C	270, I→T		Bottema et al (1991a)	516
Toronto 1 (HB31)	1	4	30,933, C→T	271, A→V		Koeberl et al (1990a)	246
HB143	<1	2	30,936, T→G	272, L→R		Bottema et al (1991b)	517

UK 48			30,942, Δ1	274	Frameshift, normal carrier female	Green, Montandon et al	247
San Antonio	<1	14	30,945, T→C	275, L→P		Jagadeeswaran	248
HB101			30,945, T→C	275, L→P		Bottema et al (1991a)	518
Malmö 1	<1	<0.1	30,950-7, Δ 8	277	Frameshift, Inhibitor	Green et al (1989)	249
Zoeterwoude	13	13	30,956, T→A	279, L→I		Reitsma et al	250
Unnamed	<1	<1	30,981, C→T	287, P→L		Chen et al (1991)	251
HB109	<1	<1	30,985, T→G	288, I→M		Bottema et al (1991a)	519
Oxford h2	2	3	30,992, G→C	291, A→P		Winship & Dragon (1991)	252
UK 13	10		30,992, G→A	291, A→T		Montandon et al (1989)	253
UK 33	7	19	30,992, G→A	291, A→T		Green et al (1992)	254
UK 41	9	11	30,992, G→A	291, A→T		Green et al (1992)	255
UK 71	7	10	30,992, G→A	291, A→T		Green, Saad et al	520
HB100	<1	<1	31,001, G→T	294, E→Stop		Bottema et al (1991a)	521
Beberbeck	4	7	31,008, C→T[4]	296, T→M		Ludwig et al	256
B.Liebenzell	6	7	31,008, C→T[4]	296, T→M		Ludwig et al	257
HB19/HB85	5	7	31,008, C→T[4]	296, T→M	Patients ID 258 & 269 are related	Koeberl et al (1989) Ketterling et al (1991)	258 269
Malmö 25	4	15	31,008, C→T[4]	296, T→M	de novo in MGF or MGM	Green et al (1991b)	259
Neuhausen	4	9	31,008, C→T[4]	296, T→M		Ludwig et al	260

UK 32	6	5	31,008, C→T[4]	296, T→M		Green et al (1990)	261
Unnamed	2	6	31,008, C→T[4]	296, T→M		Chen et al (1991)	262
Unnamed		6	31,008, C→T[4]	296, T→M		Chen et al (1991)	263
Unnamed	4	10	31,008, C→T[4]	296, T→M		Chen et al (1991)	264
Unnamed	5	11	31,008, C→T[4]	296, T→M		Chen et al (1991)	265
Farmington	2	6	31,008, C→T[4]	296, T→M		Tarnower & Smith (1991)	266
HB70	1	23	31,008, C→T[4]	296, T→M	Haplotype differs from 258	Ketterling et al (1991a)	267
HB84	6	7	31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991a)	268
HB89	5	7	31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991a)	270
Greensboro	9	8	31,008, C→T[4]	296, T→M		Rose and High	271
UK 49			31,008, C→T[4]	296, T→M		Green et al (1991b)	272
Glostrup h1			31,008, C→T[4]	296, T→M		Winship & Dragon (1991)	273
UK 73	5	10	31,008, C→T[4]	296, T→M		Green et al (1992)	274
HB21	6	6	31,008, C→T[4]	296, T→M		Koeberl et al (1990a)	393
HB22	5	8	31,008, C→T[4]	296, T→M		Koeberl et al (1990a)	394
Unnamed	3	8	31,008, C→T[4]	296, T→M		Thompson & Chen (1992)	522
Unnamed	4	10	31,008, C→T[4]	296, T→M		Thompson & Chen (1992)	523
HB126	7		31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991b)	524
Seattle M	2	11	31,008, C→T[4]	296, T→M		Thompson et al (1992)	525

Seattle N	5	17	31,008, C→T[4]	296, T→M		Thompson et al (1992)	526
HB129	15		31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991b)	527
HB138	2		31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991b)	528
HB148	2		31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991b)	529
HB167	4	4	31,008, C→T[4]	296, T→M	Same haplotype as 258	Ketterling et al (1991b)	530
Los Angeles 1	2		31,008, C→T[4]	296, T→M		Stanfield-Oakley et al	531
HB132			31,012, C→T[4]	None	Double (see 31,224) N	Bottema et al (1991b)	532
Malmö 18	<1	0.3	31,035, G→A	305, G→D	de novo in MGF	Green et al (1991b)	275
Seattle O	<1	<1	31,039, T→A	306, Y→Stop		Thompson et al (1991)	533
HB27	18	46	31,041, T→C	307, V→A		Bottema et al (1989a)	276
Unnamed	15	40	31,041, T→C	307, V→A		Chen et al (1991)	277
Seattle P	14	50	31,041, T→C	307, V→A		Thompson et al (1992)	534
Malmö 26	3	4	31,041, T→G	307, V→G		Green et al (1991b)	278
UK 122	5		31,044, G→A	308, S→N		Green, Montandon et al	279
Emsdetten	<1		31,045, T→G	308, S→R		Ludwig et al	280
Oxford h3	<1	65	31,046, G→A	309, G→S		Winship & Dragon (1991)	281
Unnamed	<1	58	31,047, G→T	309, G→V		Thompson et al (1989)	282
Toronto 21	<1	86	31,049, T→C	310, W→R		Picketts et al	283
Unnamed	<1	<1	31,051, G→A	310, W→Stop		Wang et al (1990a)	284

Albuquerque 2	<1	<1	31,051, G→A	310, W→Stop		Tarnower & Smith (1991)	285
NZ 6	<1		31,051, G→A	310, W→Stop		Van de Water et al	535
HB26	3		31,052, G→A	311, G→R		Koeberl et al (1989)	286
HB139	2	100	31,052, G→A	311, G→R		Bottema et al (1991b)	536
Amagasaki	<1	100	31,053, G→A	311, G→E	B_m[6]	Miyata et al (1991)	287
UK 137	<1	100	31,059, T→G	313, V→G		Green, Montandon et al	288
UK 11	<1	<2	31,059-60, Δ2	313	Frameshift	Green et al (1989)	289
Goldbach	<1		31,070, G→C	317, G→R		Ludwig et al	290
Toronto 7 (HB38)	<1	90	31,080, C→A	320, A→D		Koeberl et al (1990a)	291
UK 147	4		31,080, C→A	320, A→D		Green, Saad et al	537
NZ 7	3		31,080, C→A	320, A→D		Van de Water et al	538
Riegelsberg	<1	<1	31,084-90, Δ7	321	Frameshift, double (see 37), de novo in mother(30)	Ludwig et al (1992a)	14
HB82	4	4	31,091, C→T	324, Q→Stop		Bottema et al (1991a)	539
HB122	4		31,096, C→G	325, Y→Stop		Bottema et al (1991a)	540
Oxford h5	4	4	31,103, G→T	328, V→F		Winship (1990)	292
Ursem	1	<1	31,103, G→A	328, V→I	Double (see 6,491-4), N	Poort et al (1990)	77
UK 90	7		31,110, T→C	330, L→P		Green, Montandon et al	293
Ratingen 2	<1	<1	31,110-12, $\Delta 3^9$	331, ΔV	In frame	Ludwig et al	294
Oxford h1	7	96	31,113, T→C	331, V→A		Winship & Dragon (1991)	295

Brünov	4		31,115, G→T	332, D→Y		Ludwig et al	296
HB29	<1	<1	31,118, C→T[4]	333, R→Stop		Koeberl et al (1990b)	297
HB30	<1	1	31,118, C→T[4]	333, R→Stop	Haplotype differs from 297	Koeberl et al (1990b)	298
UK 34	6		31,118, C→T[4]	333, R→Stop	Female	Montandon et al (1990b)	299
Unnamed	<1	1	31,118, C→T[4]	333, R→Stop		Chen et al (1991)	300
Unnamed	<1	<1	31,118, C→T[4]	333, R→Stop		Ludwig et al	301
Unnamed	<1	<1	31,118, C→T[4]	333, R→Stop		Ludwig et al	302
Unnamed	<1	<1	31,118, C→T[4]	333, R→Stop		Ludwig et al	303
Veldhoven	<1	<1	31,118, C→T[4]	333, R→Stop		Poort et al	541
HB105	4	<1	31,118, C→T[4]	333, R→Stop	Haplotype differs from 297	Bottema et al (1991a)	542
Seattle Q	<1	<1	31,118, C→T[4]	333, R→Stop		Thompson et al (1992)	543
HB83	13	70	31,118, C→G	333, R→G		Bottema et al (1990c)	304
HB153	<1	63	31,118, C→G	333, R→G	Haplotype differs from 304	Bottema et al (1991b)	544
HB110	<1	150	31,119, G→T	333, R→L		Bottema et al (1991a)	545
UK 2	<1	135	31,119, G→A[4]	333, R→Q		Tsang et al (1988)	305
Heerde	1	122	31,119, G→A[4]	333, R→Q		Poort et al (1989a)	306
Iceland 1	3	119	31,119, G→A[4]	333, R→Q	Double (see 30,800)	Green et al (1992)	215
UK 5	1-2	75	31,119, G→A[4]	333, R→Q		Green et al (1989)	307
UK 18	1	130	31,119, G→A[4]	333, R→Q		Green et al (1990)	308

Unnamed	2	32	31,119, G→A[4]	333, R→Q		Wang et al (1990a)	309
Toronto 22	<1	87	31,119, G→A[4]	333, R→Q		Picketts et al	310
Unnamed	2	38	31,119, G→A[4]	333, R→Q		Chen et al (1991)	311
Köln	2	60	31,119, G→A[4]	333, R→Q	de novo in MGF(46)	Ludwig et al (1992a)	312
Düsseldorf	4	86	31,119, G→A[4]	333, R→Q	de novo in mother(20)	Ludwig et al (1992a)	313
Fellhammer	2	88	31,119, G→A[4]	333, R→Q		Ludwig et al	314
Bor	4	93	31,119, G→A[4]	333, R→Q		Ludwig et al	315
UK 133			31,119, G→A[4]	333, R→Q		Green, Saad et al	546
Unnamed	2	80	31,119, G→A[4]	333, R→Q		Thompson & Chen (1992)	547
HB107	2	86	31,119, G→A[4]	333, R→Q		Bottema et al (1991a)	548
Hong Kong 6	9	100	31,122, C→A	334, A→D		Chan et al (1990)	316
UK 8	2	2	31,127, T→C	336, C→R		Green et al (1989)	317
UK 139	<1		31,127, T→C	336, C→R		Green, Saad et al	549
Unnamed	<1	5	31,127, T→C	336, C→R		Chen et al (1991)	318
Malmö 24	<1	<1	31,128, G→A	336, C→Y		Green et al (1991b)	319
Calgary 3	4	120	31,130, C→A	337, L→I		Fraser et al (1992)	550
Bonn 1	<1	<1	31,133, C→T[4]	338, R→Stop		Ludwig et al (1989)	320
New York	<1	<1	31,133, C→T[4]	338, R→Stop		Driscoll et al (1989a)	321
UK 20	2	<1	31,133, C→T[4]	338, R→Stop		Green et al (1990)	322

UK 31	<1		31,133, C→T[4]	338, R→Stop		Green et al (1990)	323
Unnamed	<1	<1	31,133, C→T[4]	338, R→Stop		Freedenberg et al (1989)	324
Edmonton 2	<1	<1	31,133, C→T[4]	338, R→Stop		Tam et al	325
Calgary 2	<1	<1	31,133, C→T[4]	338, R→Stop		Poon et al (1987)	326
Calgary 4	<1	<1	31,133, C→T[4]	338, R→Stop	Double (see 48)	Poon & Sinclair	399
UK 170			31,141, 2 or 3 Ins AA	340 or 341	Frameshift, normal carrier female	Green, Saad et al	551
Samli	<1	<1	31,149-51, Δ3 31,158-62, Δ5 31,158, Ins G	343	Frameshift	Ludwig et al	327
Gladbeck	4	87	31,151, A→T	344, I→F		Ludwig et al	328
Unnamed	<1	<1	31,157, 8 or 9 Ins AA	346	Frameshift	Chen & Thompson	329
Seattle R	<1	<1	31,157, 8 or 9 Ins AA	346	Frameshift	Thompson et al (1992)	552
UK 135			31,157-9 or 31,160-2 Δ3	346 or 347 ΔN	In frame	Green, Saad et al	553
HB108	<1	2	31,161, A→T	347, N→I		Bottema et al (1991a)	554
Unnamed	3	103	31,163, A→G	348, M→V		Chen et al (1991)	330
HB124	<1	40	31,165, G→A	348, M→I		Bottema et al (1991a)	555
Offenbach	<1		31,166 or 7, Δ1	349	Frameshift, de novo in mother(29)	Ludwig et al (1992a)	331
Kingston 4	35	45	31,170, G→C	350, C→S	Somatic mosaic	Taylor et al (1991)	332

UK 17	<1	<1	31,170, G→A	350, C→Y		Green, Montandon et al	333
Malmö 47	<1		31,170, G→A	350, C→Y		Ljung et al	556
Unnamed	2	130	31,200, C→T	360, S→L		Chen et al (1991)	334
UK 144	<1		31,208, G→T	363, G→Stop		Green, Saad et al	557
Los Angeles 2	<1	110	31,208, G→A	363, G→R		Stanfield-Oakley et al	558
Eagle Rock	1-5	100	31,209, G→T	363, G→V		Bajaj et al (1990)	335
Seattle S	3	100	31,209, G→C	363, G→A		Thompson et al (1992)	559
UK 35	2	53	31,209, G→A	363, G→E		Green, Montandon et al	336
Mechtal	<1	100	31,211, G→C	364, D→H		Ludwig et al (1992b)	337
HB80	2	95	31,211, G→C	364, D→H		Bottema et al (1990c)	338
UK 30	2		31,211, G→A	364, D→N	B_m	Green, Montandon et al	339
Unnamed	<1	130	31,212, A→T	364, D→V		Chen et al (1991)	340
Varel	<1	89	31,213-14, TA→CG	365, S→G	Active site, Inhibitor, silent mutation at aa364	Ludwig et al (1988,1992b)	341
Schmallenberg	<1		31,215, G→T	365, S→I	Active site	Ludwig et al (1992b)	342
Toronto 4 (HB35)	1	90	31,216, T→A	365, S→R	Active site	Koeberl et al (1990a)	343
HB87	<1	32	31,218, G→A	366, G→E		Bottema et al (1991a)	560
Unnamed	<1	14	31,220, G→A	367, G→R		Chen et al (1991)	344
Bergamo	<1	156	31,223, C→A	368, P→T	B_m	Bertina et al (1990)	345

UK 190	3		31,224, C→T	368, P→L		Green, Saad et al	561
HB132			31,224, C→T	368, P→L	Double (see 31,012)	Bottema et al (1991b)	532
HB73	<1	<1	31,227, A→G	369, H→R		Bottema et al (1990c)	346
Malmö 16	15	13	31,248, C→A	376, T→N		Green et al (1991b)	347
Unnamed	<1	<1	31,253, T→C	378, F→L		Chen et al (1991)	348
Brantford	5		31,258, A→C	379, L→F		Tam et al	349
Barcelos	<1		31,259, A→C	380, T→P		David et al	562
Hong Kong 5	7	<10	31,260, C→G	380, T→S		Chan et al (1991)	350
UK 76	<1	1	31,260, C→T	380, T→I		Green, Montandon et al	351
Hong Kong 4	<1	<10	31,261, Δ1	380	Frameshift	Chan et al (1990)	352
Malmö 49	<1		31,276, G→A	385, W→Stop		Ljung et al	563
HB86	2	95	31,281, A→G	387, E→G		Bottema et al (1991a)	564
Mainz	<1	<1	31,287, G→A	389, C→Y		Ludwig et al	353
UK 42	4	7	31,287, G→A	389, C→Y		Green, Montandon et al	354
Lake Elsinore	<1	100	31,290, C→T	390, A→V	B_m	Spitzer et al (1988)	355
Niigata	1-4	140	31,290, C→T	390, A→V	B_m	Sugimoto et al (1988) Nishimura et al (1990)	356
Albuquerque 3	2	48	31,290, C→T	390, A→V		Tarnower & Smith (1991)	357
Unnamed	2	30	31,290, C→A	390, A→E		Wang et al (1990a) Chen et al (1991)	358

UK 151			31,290, C→A	390, A→E		Green, Saad et al	565
Angers	<1	90	31,307, G→A	396, G→R	B_m^6	Attree et al (1989)	359
Angers	<1	110	31,307, G→A	396, G→R	B_m^6	Vidaud et al (1988)	360
UK 29	<1		31,307-20, Δ14	396	Frameshift	Green, Montandon et al	361
HB10,11,12,14, 16,18	<1-6		31,311, T→C	397, I→T	Same haplotype as each other & 365-368, 379 & 380	Koeberl et al (1989) Bottema et al (1990b)	362
Long Beach	<1	100	31,311, T→C	397, I→T		Ware et al (1988)	363
Los Angeles	<1	100	31,311, T→C	397, I→T		Spitzer et al (1990a)	364
Toronto 3 (HB34)	2	55	31,311, T→C	397, I→T		Bottema et al (1990b)	365
Toronto 9 (HB40)	2	65	31,311, T→C	397, I→T		Bottema et al (1990b)	366
Toronto 10 (HB46)	1	73	31,311, T→C	397, I→T		Bottema et al (1990b)	367
Toronto 12 (HB50)	4	61	31,311, T→C	397, I→T		Bottema et al (1990b)	368
Vancouver	3	62	31,311, T→C	397, I→T		Geddes et al (1989)	369
Vancouver, Fr	2	62	31,311, T→C	397, I→T	B_m^6	Attree et al (1989)	370
Vancouver, Fr	3	70	31,311, T→C	397, I→T	B_m^6	Attree et al (1989)	371
Vancouver, Fr	4	52	31,311, T→C	397, I→T	B_m^6	Attree et al (1989)	372
Unnamed	1	45	31,311, T→C	397, I→T		Thompson et al (1990)	373
Unnamed	3	64	31,311, T→C	397, I→T		Thompson et al (1990)	374
Unnamed	2	75	31,311, T→C	397, I→T		Thompson et al (1990)	375
Unnamed	5	96	31,311, T→C	397, I→T		Thompson et al (1990)	376

Unnamed	4	56	31,311, T→C	397, I→T		Chen et al (1991)	377
Unnamed	4	88	31,311, T→C	397, I→T		Chen et al (1991)	378
HB67	<1		31,311, T→C	397, I→T		Bottema et al (1990b)	379
HB69	3	54	31,311, T→C	397, I→T		Bottema et al (1990b)	380
HB63	8	70	31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	566
Unnamed	2	72	31,311, T→C	397, I→T		Thompson & Chen (1992)	567
HB93	<1	27	31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	568
HB94	7		31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	569
HB113	8		31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	570
HB119			31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	571
HB123	3		31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	572
HB125	2		31,311, T→C	397, I→T	Same haplotype as 362	Ketterling et al (1991b)	573
HB43	8		31,326, C→T	402, S→F		Koeberl et al (1990a)	381
Lincoln Park	3	9	31,327-8, Δ2 Ins AAGGTACCAA	402	Frameshift	Rao et al (1990)	382
Unnamed	4		31,331, T→C	404, Y→H		Ludwig et al	383
Seattle T	23	26	31,334, G→T	405, V→F		Thompson et al (1992)	574
HB20	1		31,340, T→C	407, W→R		Koeberl et al (1989)	384
HB92	<1	2	31,340, T→C	407, W→R	Haplotype differs from 384	Bottema et al (1991a)	575
UK 53	<1	<1	31,342, G→A	407, W→Stop		Green, Montandon et al	385

Oxford h4	<1	0.2	31,344 or 5, Δ1	408	Frameshift	Winship & Dr[illegible] (1991)	386
HB62	<1		31,346, Ins GATT	408	Frameshift	Bottema et al (1989b)	387
Bordeaux	<1	<1	31,352, A→T	411, K→Stop		Attree et al (1989)	388
HB66			31,356, C→A	412, T→K		Bottema et al (1991a)	576

Footnotes:

1 For nucleotide numbering see Yoshitake et al (1985); for amino acid numbering Anson et al (1984).

2 Nucleotide change predicted from amino acid sequence.

3 The following comments or abbreviations are used:

(i) Inhibitor - patients developing anti-factor IX antibodies in response to therapeutic factor IX.

(ii) Frameshift - caused by the addition (symbol Ins) or deletion (symbol Δ) affecting nucleotides corresponding to the stated amino acid number and terminating at a new stop codon shortly after.

(iii) Double - a double mutant, entered twice in the data base and cross-referenced.

(iv) N - indicates the mutation, usually a double mutant, is probably a normal variant - not causing the disease.

(v) the exon (a-h) immediately adjacent to donor or acceptor splice sites is noted.

(vi) Gla refers to glutamic acid residues normally γ-carboxylated, and β-hydroxyaspartate to the single modified aspartate residue.

(vii) B_m - patients with a prolonged bovine prothrombin time (Hougie & Twomey, 1967).

(viii) de novo - refers to a new mutation originating in mother, or maternal grandmother (MGM) - mother, but not grandmother carrying the mutation, or maternal grandfather (MGF) - mother but not grandparents carrying the mutation and RFLP analysis indicating the mutant gene is inherited from the grandfather. Parental age at conception of child carrying, or affected by, the mutant gene is given in brackets, if known.

4 Indicates mutation of a CG to either TG or CA.

5 % varies with age, rising after puberty.

6 Bovine prothrombin time is moderately prolonged.

7 Patients are uniquely named, except for 'Vancouver,Fr' and 'Angers', where the authors have not distinguished different patients with the same mutation.

8 The position of insertions (Ins) corresponds to the first nucleotide of the inserted base. E.g. In Recklinghausen, 114, Ins AT, patient identity number = 17, the inserted dinucleotide AT occupies nucleotides 114 and 115 and displaces the residues normally found there, which now become residues 116 and 117.

9 Or, 31,111-3, or 31,112-4, or 31,113-5.

10 The PIN number is a patient identification number, used as an aid to patient identification now and in the future. It is not the intention that the PIN number should replace any existing nomenclature; rather, it will co-exist with it. Once a patient (and patients within the same pedigree) is given a PIN number, this will remain unaltered in updated, or new versions of this database. The reason for the slightly higher number of PINS (576) than patient entries (574) is because of amalgation of separate entries. This arose as originally separate studies are now known to refer to one and the same patient, or to patients related by pedigree.